

5

SEQUENCE LISTING

SEQ ID NO 1: Clostridium pasteurianum hydrogenase; Genbank accession number M81737
gi:144835

10 MKTHINGVQFNTDEDTTILKFARDNNIDISALCFLNNCNNDINKCEICTVEVEGTGLV
TACDTLIEDGMIINTNSDAVNEKIKSRISQLLDIHEFKCGPCNRRENCEFLKLVIKYKA
RASKPFLPKDKTEYVDERSKSLTVDRTKCLLCGRVCVNACGKNTETYAMKFLNKNKG
TIIGAEDKCFDDTNCLLCGQCIIACPVAALSEKSHMDRVKNALNAPEKHVIVAMAP
SVRASIGELFNMGFGVDVTGKIYTALRQLGFDKIFDINFGADM TIMEEATELVQRIEN
15 NGPFPMFTSCCPGWVRQAENYYPELLNNLSSAKSPQQIFGTASKTYYPSSISGLDPKNV
FTVTVMPTCKKFEADRPQMEKDGLRDIDAVITTRELAKMIKDAKIPFAKLEDSEAD
PAMGEYSGAGAIFGATGGVMEALRS AKDFAENAELEDIEYKQVRGLNGIKEAEVEI
NNNKYNVAVINGASNLFKFMKSGMINEKQYHFIEVMACHGGCVNGGGQPHVNP KD
LEKVDIKKVRASVLYNQDEHLSKRKSHENTALVKMYQNYFGKPGEGRAHEILHFKY
20 KK

SEQ ID NO 2: Desulfovibrio vulgaris hydrogenase; Genbank accession number GI:40830
X02416

25 MSRTVMERIEYEMHTPDPKADPDKLHFVQIDEAKCIGCDTCSQYCPTAAIFGEMGEP
HSIPHIEACINCGQCLTHCPENAIYEAQSWVPEVEKKLDGKVKCIAMPAPAVRYAL
GDAFGMPVGSVTTGKMLAALQKLGF AHCWDTEFTADV TIWEEGSEFVERLT KKSD
MPLPQFTSCCPGWQKYAETYYPELLPHFSTCKSPIGMNGALAKTYGAERMKYDPKQ
VYTVSIMPCIAKKYEGLRPELKSSGMRDIDATLT TRELAYMIKKAGIDFAKL PDGKR
30 DSLMGESTGGATIFGVTGGVMEALRFAYEAVTGKKPDSWDFKAVRGLDGIKEATV
NVGGTDVKVAVVHGAKRFKQVCDDVKAGKSPYHFIEYMACPGGCVC GGGQPVMP
GVLEAMDRTTTRLYAGLKKRLAMASANKA

SEQ ID NO 3: *Entamoeba histolytica* hydrogenase; Genbank accession number GI:9963974
AF248542

MPPKPSHTLTGHDHNHSIQFDWSKCMGCGMCATKCTFGVLVKQPPKIPPFVQPNRE
5 KLSQENTDKTRVLIDESECTGCGQCSLVCNFGSITPIDHLVDTFKAKEAGKKLVAMIA
PSTRLGVAEAMGMPIGSTAMAQLVHCLRLIGFDYVFDVDAGADKTTMDDYAEVIE
MKKEGKGPAITSCCPAWIELVEKEYPDLPNVSTARSPIGCLAGCIKRG
WAKDVGIAVEDLYTVGIMPCIAKKTESQRQQIHQDYDASCTSNEIAAYFKKHLPPPEE
CKFTQEREEALAKTEDGQCDLPFRRISGGSNIFGKTGGVCETVLRVIARNAGVDWNS
10 CTVNKEETFKAASGSTMTNLSVDIGGTITGAVCHGGYAIRHACELIRKGELKVDV
VEMMACVGGCLGGAGQPKIPPAKKLEMDKRRVMLDILDQQTDIRAANENTDVLGW
IDKHFDHQGAHQHLHTYFTPRYQN

15 SEQ ID NO 4: NC_001146. *Saccharomyces cerevisiae* GI:6324089

MSALLSESDLNDFISPALACVKPTQVSGGKKDNVNMNGEYEVSTEPDQLEKVSITLS
DCLACSGCITSSEEILLSSQSHSVFLKNWGLSQQQDKFLVVSVPQCRLSLAQYYGL
TLEAADLCLMNFFQKHFQCKYMGTEMGRIISISKTVEKIIAHKKQKENTGADRKPL
20 LSAVCPGFLIYTEKTKPQLVPMLLNVKSPQQITGSLIRATFESLAIARES FYHLSLMPCF
DKKLEASRPESLDDGIDCVITPREIVTMLQELNLD FKSFLTEDTSLYGRLSPPGWDPR
VHWASNLGGTCGGYAYQYVTAVQRLHPGSQMIVLEGRNSDIVEYRLLHDDR IIAAA
SELSGFRNIQNLVRKLTSGSGSERKRNITALRKRRTGPKANSREMAAATAATADPYH
SDYIEVNACPGACMNGGGLLNGEQNSLKRKQLVQTLNKRHGEELAMVDPLTLGPK
25 LEEAAARPLSLEYVFAPVKQAVEKDLSVSGSTW

SEQ ID NO 5: *Chlorella fusca* GI:21732235 AJ298227

30 MCCPVVASRHAGRARHVAVRAAGPTSECDPPTPQAKLPHWQQALDELAKPKESR
RLMIAQIASAVRVAIAETIGLAPGDVTIGQLVTGLRMLGFDYVFDTLFGADLTIM EEG
TELLHRLQDHLEQHPNKEEPLPMFTSCCPGWVAMVEKSNPELIPYLSSCKSPQMMLG
AVIKNYAQQVGVQPSDICNVSVMP CVRKQGEADREWFNTTGAGLARDVDHVVTT
AEVGKIFLERGIKLNELPESNFDNPIGEGTGGALLFGTTGGVMEALRTVYEVVTQKP

MGRVDFEVRGLEGIKEAEITLKP GDDSPFKAFAGADGGITLKI AVANGLGNAKKL
IKSLSEGKAKYDFIEVMACPGGCIGGGGQPRSTDQILQKRQQAMYNLDERSTIRRS
HDNPFIQALYDKFLGAPNSHKAHDL LHTHYVAGGIPEEK

5 SEQ ID NO 6: *Clostridium saccharobutylicum* GI:488597 U09760

MINIVIDEKTIQVQENTTVIQAALANGIDIPSLCYLNECGNVGKCGVCAVEIEGKNNL
ALACITKVEEGMVVKTNSEKVQERVKMRVATLLDKHEFKCGPCPRRENCEFLKLVI
KTKAKANKPFVVEDKSQYIDIRSKSIVIDRTKCVLCGRCEAAACKTKTGTGAISICKSES
10 GRIVQATGGKCFDDTNCLLCGQCVAACPVGALTEKTHVDRVKEAL
EDPNKHVIVAMAPSIRTSMGELFKLGYGVDVTGKLYASMRALGFDKVFDINFGADM
TIMEEATEFIERVKNNGPFTSCCPAWVRQVENYYPEFLENLSSAKSPQQIFGAAS
KTYYPQISGISAKDVFTVTIMPCTAKKFEADREEMYNEGINKNIDAVLTTRELAKMIKD
AKINFANLEDEQADPAMGEYTGAGVIFGATGGVMEALRTAKDFVEDKDLTDIEYT
15 QIRGLQGIKEATVEIGGENYNVAVINGAANLAEFMNSGKILEKNYHFIEVMACPGGC
VNGGGQPHVSAKEREKVDVRTVRASVLYNQDNLEKRKSHKNTALLNMYDYDMG
APGQGKAHELLHLKYNK

20 SEQ ID NO 7: GI:145099 M27212. *Desulfovibrio vulgaris* [Fe]-hydrogenase alpha subunit

MSRIEMEKIFYEDHAPDPKADPKLFFIQIDESKCI GCDSCQCYCPTGAIFGDTGDAH
KIPHEELCINCGQCLTHCPVGAIYESQSWVTEIEKKIKAKDVKVIAMPAPAVRYALGD
AFGLPVGTVTTGKMFSALKELGFDHCWDNEFTADVTTWEEGTEFVQRLTKKLDKPL
25 PQFTSCCPGWHKYVESLYPELFPHMSSCKSPIGMLGTLAKTYGADRMKYDRAKVYT
VSIMPCTAKKYEGMRPQLWDSGHKDIDATIDTRELA YMIKKAKIDFTKLDPDGKRDTL
MGESTGGATLFGVTGGVMEALRYAYQAVTGKKPESMDFKGV RGLQGVKEATVN
VGGVDVKVAVVHGARRFHDVCELVKAGKAPWHFIEFMACPGGCVC GGGQPVM PG
VLEAADRRSTRMYAGLKKRLAMASASRA

30

SEQ ID NO 8: GI:145100 M27212 *Desulfovibrio vulgaris* [Fe]-hydrogenase beta subunit

MQIVNLTRRGFLKAACVVTGGALISIRMTGKAVAAAKQLKDYMMDRINGVYGADA
KFPVRASQDNVQVQKLYADFLEKPM SHKAEQLLH THWVDRSKA IERMKAQGAYPN
PRAKEFEGNTYPYE

5 SEQ ID NO 9: *Desulfovibrio vulgaris* (Hildenborough) hydrogenase [gi:40827] X57838

MNAFINGKEVRCEPGR TILEAARENGHFIPTLCELADIGHAPGT CRVCLVEIWRDKEA
GPQIVTSCTTPVEEGMRIFTRTPEVRRMQRLQVELLLADHDHDC AACARHGDC ELQ
DVAQFVGLTGTRHHFPDYARSRTDVSSPSVVRDMGKCIRCLRCVAVCRNVQGV D
10 ALVVTGNGIGTEIGLRHNRSQSASDCVGGCQCTLVCPVGALAGRDDVERVIDYLYDP
EIVTVFQFAPAVRVGLGEEFGLPPGSSVEGQVPTALRLLGADVVLDTNFAADLVIME
EGTELLQRLRGGA KLPLFTSCCPGWVNF AEKHL PDILPHVSTTRSPQQCLGALAKTY
LARTMNVAPERMRVVSLMPCTAKKEEAARPEFR RDGVRDVDAVLTTREFARLLRRE
GIDLAGLEPSPCDDPLMGRATGA AVIFGTTGGVME AALRTVYHVLNGKELAPVELH
15 ALRGYENVREAVVPLGEGNGSVKVAVVHGLK AARQMVEAVLAGKADHV FVEVM
ACPGGCMDGGGQPRSKRAYNPNAQARRAALFSLDAENALRQSHNNPLIGKVYESFL
GEPCSNLSHRL LHTRYGDRKSEVAYTMRDIWHEMTLGRRVRGDSD

20 SEQ ID NO 10: *Clostridium perfringens* hydrogenase [gi:4239871] AB016775

MNKIIINDKTIEFDGDKTILDLARENGFDIPVLCELKNCGNKGQCGVCLVEQEGNDRL
LRSCAIKAKDGMVIKTDSEKVLEARKERVAELLDEHEFKCGPCKRRENCEFLKLVIK
TKARAHKPFVVADKSEYVDDRSKSIVLDRSKCVKCGRCVAACRTRTATNSIKFHRID
25 GVRLVGPEELKCFDDTNCLLCGQCIAACPVDALSEKSHIERVQEALNDPEKHVIVAM
APAVRTSMGELFKMGYGQDVTGKLYTALRELGFDK VFDINFGADMTIMEEATELIE
RIKNNGPFPMLTSCCPSWVREVENYFPELVENLSSAKSPQQIFGAASKTYYPQVADID
PKKVFTVTVM PCTSKKFEADRP EMENEGIRNIDAVITTRELARMIKAAKIDFAKLEDG
EVD PAMGEYTGAGVIFGATGGVME AALRTAKDFMENDNLDNVDYEA VRGLAGIKE
30 AEVEIAGNEYKLAVVSGAANVFELVKS GKINDYHFIEVMACPGGC VNGGGQPHISAE
DSDKMDIREVRASVLYNQDKNLEKRKSHQNSALLKMYESYMGKPGHGRAHELLH
MKYKK

SEQ ID NO 11: *Clostridium perfringens* GI:4239897 AB016820

MNKIIINDKTIEFDGDKTILDLARENGFDIPVLC ELKNCGNKGQCGVCLVEQEGNDRL
LRSCAIKAKDGMVIKTDSEKVLEARKERVAELLDEHEFKCGPCKRRENCEFLKLVIK
5 TKARAHKPFVVADKSEYVDDRSKSIVLDRSKCVKCGRCVAACTRTRTATNSIKFHRID
GVRLVGPEELKCFDDTNCLLCGQCIAACPVDALSEKSHIERVQDALNDPEKHVIVAM
APAVRTSMGELFKMGYGQDVTGKLYTALRELGFDKVFDINFGADMTIMEEATELIE
RIKNNGPFPMLTSCCPSWVREVENYFPELVENLSSAKSPQQIFGAASKTYYPQVADID
PKKVFTVTVMPTSKKFEADRP EMENEGIRNIDAVITTRELARMIKAAKIDFAKLEDG
10 EVDPAMGEYTGAGVIFGATGGVMEAALRTAKDFMENDNLDNVDYEAVRGLAGIKE
AEVEIAGNEYKLA VVSGAANVFELVKSGKINDYHFIEVMACPGGC VNGGGQPHISAE
DSDKIDIREVRASVLYNQDNLEKRKSHQNSALLKMYENYMGKPGHGRAHELLHM
KYKK

15

SEQ ID NO 12: *Megasphaera elsdenii* GI:6650985 AF120457

MPEFHSRFEKIDRRVPIDEHNCAVQFDVTKCKNCTLCRRACADTQTVLDYYSLSSTG
DMPICVHCGQCSSACPFGAIVEVNDVDKVKAAALKDPEKIVIFQTAPAVRVGLGEAFG
20 MDPGTFVEGKMVAALRTL GADYVFD TDFGADLTIMEEATELLHRLQSEEIPIQFTSC
CPAWVEFAETFYPDLLQHLSSTKSPISILSPVIKTYFAQQKNIDPKKIVNVCVTPCTAK
KAEIRRP ELSASGLFWDEPEIRDTDICITTRELAQWIQDENIDFASLEDSKFDKAFGEA
SGGGRIFGNSGGVMEAAIRTAYHMF TGRPAKDFIPFEPVRGLQGVKKATVIFGHFV
LHVAAISGLGNARAFIDDLIKNDAFEDYSFIEVMACPGGCIGGGGQPKVKLPQVKKV
25 QEARTASIYKSDEETDIKASWQNPEIETLYEAFLEPLSEMAEFTLHTYFSDKSDQLG
RMKNLTPQTNPMSPKYKPPT EE

SEQ ID NO 13: *Desulfovibrio desulfuricans* strain G20 [Fe] hydrogenase GI:13022069
30 AF331719

MNLVEMEKIQYVDQSPDPRANPDELFFIQIDPEKCIGCDTCQEYCPTGAIFGDTGSAH
SIPHEEICINCGQCLTHCPVGAIYEVQSWVRELSEKIKDPEIKVIAMPAPAVRYGLGEC
FGMPVGTVTTGKMLTALQMLGFDHVWDNEFTADV TIWEEGTEFVNRLTGQIDKPLP

QFTSCCPGWHKYVESFYPELPHLSSCKSPIGMMGALAKTYGPDVMKYDRSKVYTV
SIMPCTAKKYEGMRADLWSSGYKDIDATIDTRELAYMIKKAGIDFAALPDGKRDTL
MGDSTGGATIFGVSGGVMEALRYAYEAVTGKKPSSWDFTMVRGLNGIKEGTVTIG
DAKINVAVVHGAKRFAEVCEVIKTGKSPWHFIEFMACPGGCVCGGGQPVMPGVLEA
5 MDRKVSRTFAGLKERLNRMSSSKA

SEQ ID NO 14: U07229. Desulfovibrio fructosovorans Iron hydrogenaserogenase

GI:466366

10

MSMLTITIDGKTTSVPEGSTILDAAKTLDIDIPTLCYLNLEALSINNKAASCRVCVVEV
EGRRLAPSCATPVTDNMVVKTNLSRLVNLNARRTVLELLLSDHPKDCLVCAKSGECE
LQTLAERFGIRESFYDGGEMSHYRKDISASIIRDMDKCIMCRRCEMCNTVQTCGVLS
GVNRGFTAVVAPAFEMNLADTVCTNCGQCVAVCPTGALVEHEYIWEVVEALANPD
15 KVVIVQTAPAVRAALGEDLGVPAGTSVTGKMAAALRRLGFDHVFDTDFAAADLTIME
EGSEFLDRLGKHLAGDTNVKLPILTSCCPGWVKFFEHQFPDMLDVPSTAKSPQQMFG
AIAKTYADLLGIPREKLVVVSVMPCLAKEYECARPEFSVNGNPDVDIVITTRELAKL
VKRMNIDFAGLPDEDFDAPLGASTGAAPIFGVTGGVIEAALRTAYELATGETLKKVD
FEDVRGMDGVKKAKVKVGDNELVIGVAHGLGNARELLKPCGAGETFHAIEVMACP
20 GGCIGGGGQPYHHGDVELLKKRTQVLYAEDAGKPLRKSHENPYIIELYEKFLGKPLS
ERSHQLLHTHYFKRQRL

25 SEQ ID NO 15: Y11759 Desulfovibrio fructosovorans Iron hydrogenaserogenase

GI:1914864

MSRIEMAKIFYEQTVPPP GTNLDQAYIVQVDETKCIGCDTCMGYCPTGAITGESGEPH
KVVDPAACINCGQCLTHCPVAAIYETVSFVPEIEAKLKDKNVKVIAMPAPAVRYALG
30 DPFGMPLGAVTTEHMLTGLKQLGFDNVWDNEFTADVITWEEGSELLARITKKLDKP
LPQFTSCCPGWQKYAETFYPELLPHFSSCKSPIGMMGPLAKTYGAKELGYEPKQIYT
VSIMPCTAKKFEGMRPEMDASGFRDIDATINTRELAYMMKKAGIDLPKIANGKRDA
VMGESTGGATIFGVSGGVMEALRFAYQALTKKPPQSWDFKAVRGLNGIKEATINIG

GTDVKVAVVNGGKNFAKVCDEVKAGKSPYHFIEFMACPGGCVMGGGQPIMPTVLE
SMNRTTTTKIFYASLKKRLALYDAQKA

5 SEQ ID NO 16: *Thermotoga maritima* hydrogenase [gi:4980694] AE001705

MRRFFKNNLRNLSQNGETNSVRRCFALADVTVVINGRTLTVPDNLTVIEACEKAGIEI
PALCHHPRLGESIGACRVCVVEVEGARNLQPACVTKVRDGMVIKTSSDRVKTARKF
NLALLSEHPNDCMTCEANGRCEFQDLIYKYDVEPIFGYGTKEGLVDRSSPAIVRDLS
10 KCIKCQRCVRACSELQGMHIYSMVERGHRTPGTPDMPVYETDCIGCGQCAAF CPT
GAIVENSAVKVVLEELEKKEKILVVQTAPSVRVAIGEEFGYAPGTISTGQMVAALRR
LGFDYVFDTNFGADLTIMEEGSEFLERLEKGDLEDLPMFTSCCPGWVNLVEKVYPEL
RTRLSSAKSPQGMLSAMVKTYFAEKLGVKPEDIFHVSIMPCTAKKDEALRKQLMVN
GVPAVDVVLTTRELGKLIRMKKIPFANLP EEYDAPLGISTGAAALFGVTGGVMEAA
15 LRTAYELKTGKALPKIVFEEVRGLKGVREAEIDL DGGKIRIAVVHGTANVRNLVEKIL
RREVKYHFVEVMACPGGCIGGGGQPYSRDPEILRKRAEAIYTIDERM TLRKSHENPAI
KKLYEEYLEHPLSHKAHELLHTYYEDRSRKKRLAVK

20

SEQ ID NO 17: AE001794. *Thermotoga maritima* alpha subunit hydrogenase GI:4981990

MKIYVDGREVIINDNERNLLEALKNVGIEIPNLCYLSEASIYGACRMCLVEINGQITTS
CTLKPYEGMKVKTNTPEIYEMRRNILELILATHNRDCTTCDRNGSCKLQKYAEDFGI
25 RKIRFEALKKEHVRDESAPVVRDTSKCILCGDCVRVCEEIQGVGVIEFAKRGFESVVT
TAFDTPLIETECVLCGQCVAYCPTGALSIRNDIDKLIEALES DKIVIGMIAPAVRAAIQE
EFGIDEDVAMAEKLVSFLKTIGFDKVFDVSFGADLVAYEEAHEFYERLKKGERLPQF
TSCCPAWVKHAEHTYPQYLQNLSSVKSPQQALGTVIKKIYARKLGVPEEKIFLV SFM
PCTAKKFEAEREEHEGIVDIVLTTRELAQLIKMSRIDINRVEPQPFD RPYGVSSQAGLG
30 FGKAGGVFSCVLSVLNEEIGIEKVDVKSPEDGIRVAEVT LKDGTSFKGAVIYGLGKV
KKFLEERKDVEIIEVMACNYGCVGGGGQPYPNDSRIREHRAKVL RDTMGIKSL LTPV
ENLFLMKLYEEDLKDEHTRHEILHTTYRPRRRYPEKDVEILPVPNGEKRTVKVCLGT
SCYTKGSYEILKKLVDYVKENDMEGKIEVLGTFCVENCGASPNVIVDDKIIGGATFEK
VLEELSKNG

SEQ ID NO 18: Y16775. gi:4034790 *Nyctotherus ovalis*

5

MISRLIAKKAPLFLRTFATSEMISLKIDGKIISVPGIMLADAIKKAGANVPTMCYHPD
LPTSGGICRVCLVESAKSPGYPIISCRTPEEGMEIVTQGSKMKEYRQANLALMLSRH
PNACLSCTSNCTNCKTQELSANMNIGQCGFANATPPKNDDSYDMTTAIERDNDKCINC
DICVHTCSLQGLNALGFYNEEGHAVKSMGTLDVSECIQCGQCINRCPTGAITEKSEIR
10 PVLDAINIQQRLVFQMAPSIRVAVAEFGIKPGEKILKNEIATALRKLGSNVFLDTNF
SADLTIEEGHELIERLYRNVGTGKKLLGGDHMPIDLPMLTSCC

PGWIMFIEKNYPDLLNNLSTCKSPQGMLGALIKGYWAKNIKKMDPKDIVSVSIMPCT
AKKAEKERPQLRGDEGYKDVDYILTTRELAKMLKQSNIDLAKMEPTPFDKVMSEGT
GAAVIFGVTGGVMEAALRTANEVITGREVPFKNLNIEAVRGMEGIREAGIKLENVLD
15 KYKAFEGVTVKVAIAHGPNNARKVMDIIKQAKESGKPAPWHFVEVMACPGGCIGG
GGQPKPTNLEIRQARTQLTFKEDMDLPLRKSHDNPEIKAIYENYLKEPLGHNSHHYL
HTTYSSQKVRDMNLYNANEAAAGLDEILAKYPKEKEYLMPHIEEHDKKGYISDPSIVK
ISEHLGMYPAQIESILSSYHYFPREHTIAILMSICVHCHNCMMKGQGRLLKTIQETYDI
HETHGGVAKDGSFTLHTLNWLGYCVNDAPAMMIKRKGTNYVETFTGLLGDNIDQR
20 LKSLKNLKKELPKWPKNNIREMKSQRNGNSYSCMNTQAPIAEATKKA VSMGPEKVI
EEVFKSNLVGRGGAGFRTGKKWESAYKTPASDKYVVCNADEGLPSTYKDWCLLNN
EAKRKEVFTGMGICAKTIGAKRCFMYLRYEYRNLPALQSIKDVQSTCPELADLKY
EIRLGGGPYVAGEENAQFESIEGRAPLPRKDRPGNIFPTMEGLFHKPTVINNVETFFAI
PHIIQQGSQSFGEGKMPKLLSVTGDVDEPILNETNLNNYSLNHLLQEISAKDIVAAEIG
25 GCTEPIIFGSKFDTLFGFGRGTLNAVGSVVLFNSSCDLGKIYENKLKFM AEESCKQCV
PCRDGSYIFHRAFKELRDTGKSSYNMRALAVASESAARSSICAHGKALES LFKSACDF
MNKTKPIYQPHSTYHQ

30

SEQ ID NO 19: AF262402 *Spironucleus barkhanus* GI:11127703

10 MKVRQSPFKIDITNGPIDRND AIQIDYQKCIGCQMCAKTCTDSQNFNIFKISAPKTKPF
VNAYGSVAEGTERNALAGTDCTGCGACVRACVREALMPAFNIRPVLEPISEKKKVTI
AVIAPSTRVGLAEGMGMGVGVTAERQM VYELKQMGFDYVFDNMWGADAPTTEDA
KEILKAKAAGKTAFTSCCPAWVKLVETTYPELLPNISSARSPHGIICSVIKKYFAKDIG
5 KKADELYVVGVMPTAKKNEAARKELTTDGSPDCDISITREL MAYLKEKKVTFSA
AREIELKDNVQAQYDAPFNTFSGSAYTYGKTAGVTEAVVRYVCAIKKVPFSVGMITK
ELIWENKLHSSSLTLLTFSAAGEDYRICVSYGGLAHKAVELYKSGELKVD AVEVM
VCPGGCVGGGGQPKQPKKDMILKRHEGLDKHDKEAPYSNCTENPTLNEFYERIGTD
VHHVMHTTYSAYK.

10

SEQ ID NO 20: U19897. *Trichomonas vaginalis* gi:1171116

15 MLASSATAMKGFANSLRMKDYSSTGINFDMTKCINCQSCVRACTNIAGQNVLKSLT
VNGKSVVQTVTGKPLAETNCISCGQCTLGCPKFTIFEADAINPVKEVLTKKN GRIAVC
QIAPAIRINMAEALGVPAGTISLGKVVTALKRLGFDYVFDTNFAADMTIVEEATELVQ
RLSDKNAVLPMTSCCPAWVNYVEKSDPSLIPYLSSCRSPMSMLSSVIKNVFPKKIGT
TADKIYNVAIMPCTRKKDEIQRSQFTMKDGKQETGAVLTSRELAKMIKEAKINFKEL
PDTPCDNFYSEASGGGAIFCATGGVMEAAVRSAYKFLTKKELAPIDLQDVRGVASG
20 VKLAEVDIAGTKVKVAVAHGIKNAMTLIKKIKSGEEQFKDVKFVEVMACPGGC VVG
GGSPKAKTKKAVQARLNATYSIDKSSKHRTSQDNPQLLQLYKESFEGKFGGHVAHH
LLHTHYKNRKVNP

25

SEQ ID NO 21: U26964. *Trichomonas vaginalis* gi:1345093

30 MLASSSRAAANIRWVDTSHNAIAFDMHKCINCQACVRACKNVAGQSVLKS VKINEG
KKKGVVQTVTGKLLAETNCIGCGQCTLVCPTQAIHEKDALKQMNNIFKNKGDRILV
CQIAPAIRINMRRPWCSSRNSFHRQSRYSQRLGFDYVFDTNFGADLTIVEEATELLQ
RLNDPKAVLPMTSCCPAWVNYVEKSYPQWMPHLSTCRSPIGMLSAVIKNVFPKHIG
VDPKRIFSVGIMPCTAKKDEAAREQLMTKSGLHETDL DITSRELAKMIKAAKINFKEL
PDTELDSPYAMATGGGAIFCATGGVMEAAVRSAYKFATGKELAPIEFVQVRGAEKGI

KVGTVDINGREIKVAVAQGVKNAMSLIKKIEEGQDDVKGVVFCEVMACPGGCVGG
GGSPRAKTKAAMNKRLDATYRIDRASKYRTPQDNTQLQDLYNATWVVS L VMD

5

SEQ ID NO 22: AF262401. *Trichomonas vaginalis* gi:11127700

ASTGINSTANILRNITVTVNGKPLEAKKGETVLELCDRNNIRIPRLCFHPNLPPKASCR
VCLVECDGKWLSPACVTTVWDGLKIDTKSKNVRDSVENNLKELLDCHDETCSACIA
10 NHRCQFRDMNVAYSVKAETKEICSEEGIDESTNAIRLDTSKCVLCGRCIRACEEVAGT
SAIIFGNRAKKMRIQPTFGVTLQETSCIKCGQCTLYCPVGAITEKSQVKEALDILANKG
KKITVVQVAPAVRVALSEAFGYKEGTVTTGKMVSALKALGFDLVYDTNYGADLTIC
EEAGELVNRLRDPNAKFFMFTTCCPAWVNYVEQSAPDFIPNLSSCRSPQGMLSALIK
NYLPKLLDVKQEDVLNFSIMPCTAKKDEVERPELRTKSGLKETDMVLTVREL VEMIK
15 LSNIDFNNLPDTQFDNIFGFGSGAGQIFAATGGVMEAASRTAFEVYTGKKLTNVNIYP
VRGMDGLRIAELDLDTGTLKLVAVCHGIAN TAKLLDRLREKDPELMDIKFIEIMACPG
GCVCGGGTPQPKNRVSLDNRLAAIYNIDAKMECRKSHENPLIKGVYKEFLGKPNSHL
AHELLHTHFKHHPKW

20

SEQ ID NO 23: U15277 *Trichomonas vaginalis* gi:557063

25 MKTIILNGNEVHTDKDITILELARENNDIPTLCFLKDCGNFGKCGVCMVEVEGKGF
RAACVAKVEDGMVINTESDEVKERIKKRVSMMLDKHEFKCGQCSRRENCEFLKLVI
KTKAKASKPFLPEDKDALVDNRSKAIVIDRSKCVLCGRCVAACKQHTSTCSIQFIKKD
GQRAVGTVDDVCLDDSTCLLCGQCVIACPVAALKEKSHIEKVQEALNDPKKHVIVA
MAPSVRTAMGELFKMGYGKDVTGKLYTALRMLGFDKVFDINFGADMTIMEEATEL
30 LGRVKNNGPFPMTSCCPAWVRLAQNYHPELLDNLSSAKSPQQIFGTASKTYYPSSISG
IAPEDVYTVTIMPCNDKKYEADIPFMETNSLRDIDASLTRELAKMIKDAKIKFADLE
DGEVDPAMGTYSGAGAIFGATGGVMEAAIRSAKDF AENKELENVDYTEVRGFKGIK
EAEVEIAGNKLNVA VINGASNFFEFMKSGKMNEKQYHFIEVMACPGGCINGGGQPH

VNALDRENVDRKLRASVLYNQDKNVLSKRKSHDNPAIIKMYDSYFGKPGEGLAH
KLLHVKYTKDKNVSKHE

5 SEQ ID NO 24: AF289201 *Chlamydomonas reinhardtii* gi:9837539

MSALVLKPCAAVSIRGSSCRARQVAPRAPLAASTVRVALATLEAPARRLGNVACAA
AAPAAEAPLSHVQQALAEAKPKDDPTRKHVCVQVAPAVRVAIAETLGLAPGATTP
KQLAEGLRRLGFDEVFDTLFGADLTIMEEGSELLHRLTEHLEAHPHSDEPLPMFTSCC
10 PGWIAMLEKSYPDLPYVSSCKSPQMMLAAMVKSYLEKKGIAPKDMVMVSIMPCT
RKQSEADRDFWCVDADPTLRQLDHVITTVELGNIFKERGINLAELPEGEDWNPMGV
GSGAGVLFGTGGVMEALRTAYELFTGTPLPRLSLSEVRGMDGIKETNITMVPAPG
SKFEELLKHRAAARAEAAAHGTPGPLAWDGGAGFTSEDGRGGITLRVAVANGLGN
AKKLITKMQAGEAKYDFVEIMACPAGCVGGGGQPRSTDKAITQKRQAALYNLDEKS
15 TLRRSHENPSIRELYDTYLGEPLGHKAHELLHHTHYVAGGVEEKDEKK

SEQ ID NO 25: AJ298228 *Chlorella fusca* GI:18073435

20 AGPTSECDPPTPQAKLPHWQQALDELAKPKESRRLMIAQIASAVRVAIAETIGLAPG
DVTIGQLVTGLRMLGFDYVFDLFGADLTIMEEGTELLHRLQDHLEQHPNKEEPLPM
FTSCCPGWVAMVEKSNPELIPYLSSCKSPQMMLGAVIKNYAQQVGVQPSDICNVS
VMPCVRKQGEADREWFNTTGAGLARDVDHVTTAEVGKIFLERGIKLNELPESNFD
NPIGEGTGGALLFGTTGGVMEALRTVYEVVTQKPMGRVDFEEVRGLEGIKEAEITL
25 KPGDDSPFKAFAGADGQGITLKIAVANGLGNAKKLIKSLSEGKAKYDFIEVMACPGG
CIGGGGQPRSTDQILQKRQQAMYNLDERSTIRSHDNPFIQALYDKFLGAPNSHKA
HDLLHHTHYVAGGIPEEK

30

SEQ ID NO 26: AY055756. *Chlamydomonas reinhardtii* GI:18026272

MALGLLAELRAGQAVACARRTNAPAHAAVVPCLPSRAGKFFNLSQKVPSSQSARG
STIRVAATATDAVPHWKLALAEELDKPKDGGKVLIAQVAPAVRVAIAESFGLAPGA

VSPGKLATGLRALGFDQVFDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMF
TSCCPGWVAMMEKSYPELIPFVSSCKSPQMMMAMVKTYLSEKQGIPAKDIVMVS
MPCVRKQGEADREWFCVSEPGVRDVDHVITTAELGNIFKERGINLPELPDSDWDQPL
GLGSGAGVLF GTTGGVMEALRTAYEIVTKEPLRLNLSEVRGLDGIKEASVTLVPA
5 PGSKFAELVAERLAHKVEEAAAAEAAAAVEGAVKPPIAYDGGQGFSTDDGKGGLKL
RVAVANGLGNAKKLIGKMVSGEAKYDFVEIMACPAGCVGGGGQPRSTDQITQKR
QAALYDLDERNTLRRSHENEAVNQLYKEFLGEPLSHRAHELLHTHYVPGGAEDA

10 SEQ ID NO 27: AF276706. *Scenedesmus obliquus* GI:12581498

PHWQQTLDELAKPKERKVMIAQIAPAVRGIAETMGLNPGDVTVGQMVTGLRMLGF
DYVFDTLFGADLTIMEEGTELLHRLQDHLEQHNPKEEPLPMFTSCCPGWVAMVEKS
NPELIPYLSSCKSPQMMLGAVIKNYFAAEAGAKPEDICNVSVMPVCVRKSGEAEPRSG
15 STHHRAGR RDVDHVMTTAELGKIFVERGIKLNELQESPFDPNVGEGSGGGLLFGTTG
GVMEALRTVYEVVTAELGPQRSSLTTSTAWTPAQRASPRSPQAPTAPSRPLQAQ
TESGITLNI AVANGLGNAKKLIKQLAAGESKYDFTEVMACPGGCIGGGGQPQRNKQI
LQKRQAAMYDLDERA VIRRTENPLIGALYEKFLGEPNGHKAHELLHTHYVAGGVPD
RRSEGEAW

20

SEQ ID NO 28: NC_003869 *Thermoanaerobacter tengcongensis* strain MB4T gi:20806542

25 MDKVRVTIDGITVEVPSYYTVLEAAKEAGIDIPTLCYLKEINQIGACRICLVEIEGVRN
LQTSCTYPVFDGMKVYTNTPKIREARRLNLELILSNHDRNCLTCVRSTNCELQALAK
RLGV EIRFEGENIKYPIDDA SPAVVRDPNKC VLCRRCVAVCSEVQNVFAIGMVNRG
FKTMVAPSFGRSLKDSPCISCGQCIMVCPVGAIYEKDHTKRVYEALADDKKYVVAQ
TAPAVRVALGEEFGMPVGTIVTGKMAAALRRMGF DAVFDTNFAADLTIMEEGSELL
30 ERIKHGGKLP MITSCSPGWIAFCEKYYPEFIDNLSTCKSPHMMMGALVKSYYAEKKG
LDPKDIFVVSIMPCTAKKLEIEREEMIRNGMKD VDAVLT TRELARMIKEMGIDFVNL
KDEEFDEPLGMSTGAGAI FGATGGVMEALRTVAEIVEGRDIGKIDFEEVRGLEGVR
EATITIDGMDIKIAIANGTGN AKKLLDKVKAGEVEYHFIEVMGCPGGCIMGGGQPIH

NPNEMEEVKKLRAKAIYEIDKNLPIRKSHENPAIKRLYEFLGYPLSEKSHELLHTHY
SRKELYPLVK

5

SEQ ID NO 29: AY033895. *Neocallimastix frontalis* hydrogenase gi:19547862

MSMLSSVLNKA VVNPKLTRSLATAAAEKMVNISINGRKFQVKPKTTVLEAAKANGY
YIPTLCYHQELPVAGNCRLCLVYAKGSWKPLTACTTEVWEGMEIETDSPAVIETVRS
10 SLSMMREEHPNDCMTCGSNGDCEFQDLIYRYQIDAKHPVRSLLKHKS KKTNHSITEP
CYSPFDNTTFSVARDMNKCVKCGRCIRACHHFQNNILGFINRAGYERVGTPMDRPM
NFTKCVCEGQCSQVCPVGAITARTEVVDVLRHLDTKRKVVVCSTAPAIRVAPAEFFS
TEADFDFTGKMVAGLRKLGFDYIFDTNFSADLTIMEEGTELIDRLNNGGKFPMFTSC
CPGWINMVEKSYPELSDNLSSCKSPQQMIGAVIKSYFAKKLGLSTEDIIHVSIMPCTA
15 KKGEARRPEFVQKGKDGDYDPDIDYVITTRELLTLLKLKKINPAELPDDKFDSPLGIG
SSAGNLFGVTGGVMEAAIRTAQVITGVENPIPLGELKAIRGLDGIKAANVPLKTKDG
KEVSVRAAVVSGGANIQKFLEKIKNKELEFDFIEMMMCPGGCINGGGQPKSADPEIV
AKKMQRMYTMDDQAKLRLCHENPEIIDVYKNFLGEPNSHLAHELLHTHYNDRSKI
HDMGHHEKK

20

SEQ ID NO 30: AF446076. *Piromyces* sp. E2 GI:19548180

25 CLVDVKGSWKPLTACTTEVWEGMEIETDTPAVRETVRSSLAMMREEHPNDCMTCS
NGNCEFQDLIYRYQIDAQHPVRTLLRNKFKKTNHSITEPCYSPFDDSTFSISRDMNKC
VKCGRCVRACHHFQNNILGFINRAGYERVGTPMDRPMNFTKCVCEGQCSQVCPVG
AITERNECIEVLRHLDTKRKIVVVSTAPAIRVALAEEFNAEPDFDFTGKMVAGLKKLG
FDYIFDTNFSADLTIMEEGTELITRLNEGGKFPMFTSCCPGWINMVEKSYPEIRDNLSS
30 CKSPQQMIGAVIKTYFAKKINAKPEDIIHVSVMPTAKKGEAKRPEFKRDGVPDIDHV
ITTRELITLLKLKRINPSELKNEKFDSPLGIGSSAGNLFGVTGGVMEAAVRTAQITGV
ENPIPLGELKAIRGLDGIKAASVPLKTKDGKDVNVRAAVVSGGANIQKFLEKLKKKE
LEFDFVEMMMCPGGCINGGGQPKSADPKVVAKKMERMVYTMDDQASLRLSHENPEI
TQIYKEFLKEPNGHLSHELLHTHYNDRSKAIQDMSLHQK

SEQ ID NO 31: AF516683. *Neocallimastix frontalis* GI:23664246

5 TERNEVIEVLRQLDSKRKILVCSTAPAIRVALAEFNADPDFNFTGKMVAGLRKLG
DYIFDTNFSADLTIMEEGTELINRLNNGGKFPMFTSCCPGWINMVEKSYPELRENST
CKSPQQMIGALIKSYFAKKLGVSTEDIIHVSVMPTAKKGAEKRPEFVQKGKDGKNY
PDIDYVLTRELLTLMKLKKVNP AELADDKLDSPLGISSSAGNLFVGTGGVMEAAVR
TAQITGVENPIPLGELKAVRGLEGIKAATVPLKTKEGKDINVRAAVVSGGANIQKFL
10 EKIKNKEVEFDFVEMMMCPGGCINGGGQPKSADPKIVTKKMQRMYTMDEQATLRL
SHENEEVKQIYKEFLIEPNGHLSHELLHTHYNDRSKAIQDMSLHEKK

15 SEQ ID NO 32: NZ_AABI01000003. *Desulfovibrio desulfuricans* G20 GI:23474435

MNGQQNVIRIDSDICTGCGRCKDVCPVGAVEGVQGTPHSIREDCVLCGQCVCQQCS
AFASFYEQHPACIAEKKRERGLFVSEAAPLFAAWHTGDAPRVAGRLAEGCHSMVQC
APAVRAAIGEEFGMPAGALTPGRLAAALRRLGFDRVYDTNFAADLTIMEEGSELLQR
20 MEGAGPLPMFTSCCPAWVRYAEQQFPDLLEHLSSCKSPQQMAGAVFKSYGAQLDG
VDPRQVFSVAVMPCTCKKAEAQRPMEHDGVRDVEDAVLTTGELAAMLRQAHIDFA
ALPDEPFDRPLGSYSYSGAGNIFGLTGGVMEAAALRTAYELVTGEPVPCTELVYVRGGEG
IRHATLTMDGRTRFVAVVAGLQHVRPLLEAVRAGTCDVNFVEVMCCPQGCISGGGQ
PKVLLPFQRDEVYAARKAALYRHDAELACRKSHENPQVQALYREFLGEPLSHVSHN
25 LLHTVYGQTR

SEQ ID NO 33: NZ_AABB01000250. *Desulfitobacterium hafniense* GI:23112964

30 MMQLKHPFQSGFQQQSCKRHTKKVVVDMESKAGKGSNLSRRSFLKFAGGAGIAGA
SLSLTGCGQPLTPASAVGGEGWMPTQYNEPGGWPTNVRGRVPIDPENPALRRDDQK
CILCGQCIEVCKTIQSVYGNIELPLKNEIPCINCGQCIHWCPSGAISEREDIDQVAKAL
ADPKITVVVQTAPATRIGLGEEFGLPVGTVNQGKQVAALRKLGFDFVIF

DTNFAADLTIMEEGTELVKRITGELHHPLPQFTSCCPGWVKFVEYYYPELLPNLSSAK
SPQQMAGALVKTYFAEKNHVEPQKIFSVAIMPCTAKKFECQRPEMISAQTYWQDEQ
VSPDVDVVLTTRELARMIKRAGIDLPSLPDEEYDQLMGVATGAGAIFGTTGGVMEA
AVRSAYYLVTGEQPPAALWQLTPVRGMEGVKEAAVSIPGAGEIRIAVISGLDNARAI
5 MEQVKAGNSPWTFIEVMACPGGCQYGGGQPRSSAPPSDGVNRNTRAASLYKIDAQAK
LRNSHDNPQIKQVYAEFLTSPLESEKAEELLHTHYISRAEEFDAKKPQSHEYEV

SEQ ID NO 34: AJ312124. *Eubacterium acidaminophilum* GI:14250935

10 MVNITIDGRQVTPANSTVLDAARDMGINIPTLCYLKDINKTGACRMCLVEVEGIRN
LQTACTFPVRDGLVVKTNTRKVRDARRDNLQLILSNHHRDCLSCFRNGSCELQALC
DDMGLSELDFEAPKELKPVDMLSHSIVRDPNKCILCGRCVAVCNKVQEVGILAFTR
GVETEVAFAFATSMADAPCIYCGQCVNVCPVAALREKTDIEKVWEVLEDETKHVVV
15 QVAPAVRAALGEMFGNPIGTRVTGKMFTALKMLGFQKVFDTNFAADLTIMEEGTEL
LGRIKNGGTLPMITSCSPGWIRYVEHFYPELLDHVSSCKSPQQMMGAVLKSYYAEKN
NIAPENMIVVSVMPICIAKKTESAKEEMKNVHGTRDVIDVLTTRELGKMIKEARIEFN
DLQDSNPDEFFGDYTGAAVIFGATGGVMEAAIRTVADIVSGQELEDIEYTAVRGLEGI
KEAAVKIGDLEVKVAVAHGTANAGKLM DLVRD GKADYHFIEIMGCSGGCVTGGGQ
20 PHVDSRTKEKVNVKLERAKALYTEDKLRDKRKSHHNESVKRLYEEYLGKPNGHKA
HELLHTHYKKRELF

25 SEQ ID NO 35: NZ_AAAF01000001. *Rhodopseudomonas palustris* GI:22962132

MCTPDQASLSARDPAEATITLSINGVACAGFANETILSCARRYDVYIPTLCELEDIDHT
PGACRVCLVEILQAGKDTPIVTACNTPVRDGMVQTRSKKARDMQRLQVELLMA
DHLQDCATCIRHGSCELQDLAQFVGLQQNRFFDRERTEARPVDHSSPSMVRDMRRC
30 VRCQRCVAICRYHQKIDALAIEGSGLERMVALRDADGYPNVSCVSCGQCVLVCPTG
ALGERDETRALDYICDPNVVTVVQFAPAVRVAFGEEFGLPAGTNVEGQIIAACRKL
GVDVVLDTNFAADVIMEEGAELLARLKQGRRPTFTSCCPAWINFAEIHYPDVLPLL
SSTKSPQQVLSTIAKSYLPAQLGVPAERIRVISIMPCIAKKDEAVRPQM VHDGQPETD
LVLTTREFARLLRREGIDLKDLPSQFDRPFLSAYSGAGAIFGTTGGVMEAAVRTIYA

LVNGRELERIELTQLRGFEGLEATVDLGAPVGEVKVAMVHGLGDTRKLVESVLSG
EANYDFIEVMACPGGCVDGGGSLRSKKAYLPLALKRRETTYNVDRAAKVRQSHNNP
QVQALYRELLQAPNSEIAHRLHHTHYASRKRELQHTVKEIWDDLTMSTILY

5

SEQ ID NO 36: NZ_AAAJ01000020. Clostridium thermocellum GI:23021053

MDSFLMKGYIKEANIDYSCSRGSMEDLPKWEFREIPKVPRAVMPSLSLEERKNNFNE
10 VELGLSEEVARKEARRCLKCGCSARFTCDLRKEASNHGIVYEEPIHDRPYIPKVDDHP
FIVRDHNKCISCGRCAACAEIEGPGVLTfYMKNGRQLVGTKSGLPLRDTDCVSCGQ
CVTACPCAALDYRRERGKVVRANDPKKTVVGFVAPAVRSLISNTFGVSYEEASPFM
AGLLKKLGFDKVFDFTFAADLTIVEETTEFLSRIQNGVMPQFTSCCPGWINFVEKRY
PEIHPHLSTCKSPQMMMGA TVKNHYAKLMGINKEDLFVVSIVPCLAKKYEAAARPEFI
15 HDGIRDVDAVLT TTTEMLEMMELADIKPSEVVPQEFDEPYKQVSGAGILFGASGGVAE
AALRMAVEKLTGKVLTDHLEFEEIRGFEGVKESTIDVNGTKVRVA VVSGLKNAEPIIE
KILNGVDVGYDLIEVMACPGGCICGAGHPVPEKIDSLEKRQQVLVNIDKVS KYRKSQ
ENPDILRLYNEFYGEPNSPLAHELLHTHYTPKHGDSTCSPERKKGTAAFDVQEFTICM
CESCMEKGAENLYNDLSSKIRLFKMDPFVQIKRIRLKETHPGKGVYIALNGKQIEEPM
20 LSGNIPDESESE

SEQ ID NO 37: AB035092 Clostridium perfringens GI:7959055

MNKIIINDKTIEFDGDKTILDLARENGFDIPVLC ELKNCGNKGQC GVCLVEQEGNDRL
25 LRSCAIKAKDGMVIKTDSEKVLEARKERVAELLDEHEFKCGPCKRRENCEFLKLVK
TKARAHKPFV VADKSEYVDDRKSIVLDRSKCVKCGRCVAACRTRTATNSIKFHRID
GVRLVGPEELKCFDDTNCLLCGQCIAACPVDALSEKSHIERVQEALNDPEKHVIVAM
APAVRTSMGELFKMGYGQDVTGKLYTALRELGFDKVFDINFGADMTIMEEATELIE
30 RIKNNGPFPMLTSCCPSWVREVENYFPELVENLSSAKSPQQIFGAASKTYYPQVADID
PKKVFTVTVMPC TSKKFEADRP EMENEGIRNIDAVITTRELARMIKAAKIDFAKLEDG
EVDPA MGEYTGAGVIFGATGGVMEAAALRTAKDFMENDNLDNVDYEAVRGLAGIKE
AEVEIAGNEYKLAVVSGAANVFELVKSGKINDYHFIEVMACPGGC VNGGGQPHISAE

DSDKMDIREVRASVLYNQDKNLEKRKSHQNSALLKMYESYMGKPGHGRAHELLH
MKYKK

5

SEQ ID NO 38: NZ_AAAJ01000005. *Clostridium thermocellum* GI:23020350

MHVLKLVHSTQYWRAEEMDNREYMLIDGIPVEINGEKNLLELIRKAGIKLPTFCYHS
ELSVYGACRMCMVENEWGGLDAACSTPPRAGMSIKTNTERLQKYRKMLELLLAN
10 HCRDCTTCNNNGKCKLQDLAMRYNISHIRFPNTASNPVDVDDSSLCITRDRSKCILCGD
CVRVCNEVQNVGAIDFAYRGSKMTISTVFDKPIFESNCVGGCQCALACPTGAIVVKD
DTQKVWKEIYDKNTRVSVQIAPAVRVALGKELGLNDGENAIGKIVAALRRMGFDDI
FDTSTGADLTVLEESAELLRRIREGKNDMPLFTSCCPAWVNYCEKFYPELLPHVSTCR
SPMQMFASIIKEEYSTSSKRLVHVAVMPCTAKKFEAARKEFKVNGVNPVDYVLTTQ
15 ELVRMIKESGIVFSELEPEAIDMPFGTYTGAGVIFGVSGGVTEAVLRRVSDKSPTSFR
SLAYTGVRGMNGVKEASVMYGDRKLKVAVVSGLKNAGDLIERIKAGEHYDLVEV
MACPGGCINGGGQPFVQSEEREKRGKGLYSADKLCNIKSSEENPLMMTLYKGILKGR
VHELLHVDYASKKEAK

20

SEQ ID NO 39: NZ_AABI01000004. *Desulfovibrio desulfuricans* GI:23474512

MAGCKAQHPPAAYLAGLEVPAAGSEVTMEGVRYKMNAPKDVPATIRFVEVDHDK
CMACGECEYHCPTGVMQEVTEDEGYRGVVDPVACVNCGQCLANCPFGAIHEEVSVFV
25 GELYEKLKDPDTVVVSMPAPAVRYALGECFGLPTGTYYVGGQMHAALRRLGFNLVW
DTEWTADVTIMEEGTELLERVKHGNNMPLPQFTSCCPGWIKFAETFYPDLEKHLSTCK
SPIAMIGPLAKTYGAQEAGVPAKKMYTVSIMPICIAKKFEGMRPEMNASGYRDIDATI
TTRELAWMIKKAGIDFTSLPSEEPDPAALGMSTGAATIFCTSGGVMEAAALRLAYEALS
GGTLADPDIKVVRTHEGINTAEVVPNFGTVKVAVASGLDNAAKLCEEVRAGKSPY
30 HFIEVMTCPGGCVNGGGQPLEPGMLQSSLFKSTITTKINRRFTRRSVA

SEQ ID NO 40: NZ_AABI01000049. *Desulfovibrio desulfuricans*. GI:23475985

MNLVEMEKIQYVDQSPDPRANPDELFFIQIDPEKCIGCDTCQEYCPTGAIFGDTGSAH
SIPHEEICINCGQCLTHCPVGAIYEVQSWVRELSEKIKDPEIKVIAMPAPAVRYGLGEC
FGMPVGTVTTGKMLTALQMLGFDHVWDNEFTADVTIWEEGTEFVKRLTGQIDKPLP
5 QFTSCCPGWHKYVESFYPELFPHLSSCKSPIGMMGALAKTYGPDVMKYDRSKVYTV
SIMPCTAKKYEGMRADLWSSGYKDIDATIDTRELAYMIKKAGIDFAALPDGKRDTL
MGDSTGGATIFGVSGGVMEAAALRYAYEAVTGKKPSSWDFTMVRGLNGIKEGTVTIG
DAKINVAVVHGAKRFAEVCEVIKTGKSPCISSSLCLPR

10

SEQ ID NO 41: NZ_AABI01000001. Desulfovibrio desulfuricans GI:23473140

MNLVEMEKIQYVDQSPDPRANPDELFFIQIDPEKCIGCDTCQEYCPTGAIFGDTGSAH
15 SIPHEEICINCGQCLTHCPVGAIYEVQSWVRELSEKIKDPEIKVIAMPAPAVRYGLGEC
FGMPVGTVTTGKMLTALQMLGFDHVWDNEFTADVTIWEEGTEFVKRLTGQIDKPLP
QFTSCCPGWHKYVESFYPELFPHLSSCKSPIGMMGALAKTYGPDVMKYDRSKVYTV
SIMPCTAKKYEGMRADLWSSGYKDIDATIDTRELAYMIKKAGIDFAALPDGKRDTL
MGDSTGGATIFGVSGGVMEAAALRYAYEAVTGKKPSSWDFTMVRGLNGIKEGTVTIG
20 DAKINVAVVHGAKRFAEVCEVIKTGKSPWHFIEFMACPGGCVCVGGGQPVMPGVLEA
MDRKVSRTFAGLKERLNRMSSSKA

25 SEQ ID NO 42: AY028641. Trichomonas vaginalis GI:19547861

CDGKWLS PACVTTVWDGLKIDTKSKNVRDSVENNLKELLDCHDETCSACIANHRCQ
FRDMNVAYS VKAETKEICSEEGIDESTNAIRLDTSKCVLCGR CIRACEEVAGTSAIFG
NRAKKMRIQPTFGVTLQETSCIKCGQCTLYCPVGAITEKSQVKEALDILANKGKKITV
30 VQVAPAVRVALSEAFGYKEGTVTTGKMVSALKALGFDLVYDTNYGADLTICEEAGE
LVNRLRDPNAKFPMFTSCCPAWVNYVEQSAPDFIPNLSSCRSPQGMLSALIKNYLPK
LLDVKQEDVLNFSIMPCTAKKDEVERPELRTKSGPKETDMVLTVRELVEMIKLSNID
FNNLPDTQFDNIFGFGSGAGQIFAAT

SEQ ID NO 43: AF446077. *Trichomonas gallinae* GI:19548182

CDGKWLSPACVTTVWDGLRIDTKSKVVRDSVENNLKELLDCHDETCSSCVANHRC
5 QFRDMNVAYSVKADTKEICSEEGIDESTHAIRLDTSKCVLCGRCIRACEEVAGTSAIIF
GNRAKHMRIQPTFGGTLQETACIKCGQCTLYCPVGAITEKSQVKEALDILANKGKKV
TVVQVAPAVRVALSEAFGYKEGTVTTGKMVSALKALGFDLVYDTNYGADLTICEEA
GELVNRLKDPKAVFPMFTSCCPAWVNYVEQSAPDFIPNLSSCRSPQGMLSSLIKNYLP
KLLGIKQEEVMNFSIMPCTAKKDEIERPELQTKTGLKETDMVLTVRELVEMIKLSNID
10 FNNLPDTPFDNIFGFGSGAGQIFAAT

SEQ ID NO 44: Y16669 *Nyctotherus ovalis* GI:4029298

15 MISRLIAKKAPLFLRTFATSEMISLKIDGKIISVPKGIMLADAIKKAGANVPTMCYHPD
LPTSGGICRVCLVESAKSPGYPIISCRTPVEEGMEIVTQGSKMKEYRQANLALMLSRH
PNACLSCTSNTNCKTQELSANMNIGQCGFANATPPKNDDSYDMTTAIERDNDKCINC
DICVHTCSLQGLNALGFYNEEGHAVKSMGTLDVSECIQCGQCINRCPTGAITEKSEIR
20 PVLDAINIQQRLVFQMAPSIRVAVAEFGIKPGEKILKNEIATALRKLGSNVFVLDTNF
SADLTIIIEGHELIERLYRNVGTGKKLLGGDHMPIDLPMLTSCCPGWIMFIEKNYPDLL
NNLSTCKSPQGM LGALIKGYWAKNIKKMDPKDIVSVSIMPCTAKKAEKERPQLRGD
EGYKDVDYILTTRELAKMLKQSNIDLAKMEPTPFDKVMSEGTGAAVIFGVT

25

SEQ ID NO 45: AY028640 *Trichomonas vaginalis*. gi:19547858

CDGKWLAPACVTTVWDGLKIDTKSKMVKESVENNLKELLDCHDETCSSCVANHRC
QFRDMNVAYSIIKAETKEECSEEGIDESTNSIRLDTSKCVLCGRCIRACEEVAGQSAIIF
30 GNRAKHMRIQPTFGQTLQDTSCIKCGQCTLYCPVGAITEKSQVKQALDILSNKGKKIS
VIQVAPAVRVALSEAFGYKEGTVTTGKMVSALKALGFDYVYDTNYS
ADLTIVEEAGELVQRLKNPNAVFPMTSCCPAWVNYVEQSAPDFIPNLSSCRSPQGM
LSSLVKNYLPKVLNIPVEDVLNFSIMPCTAKKDEIERPELRTKDGHKETDMVLTVREL
VEMIKLSGIDFNNLPDTPFDSIFGFGSGAGQIFAAT

SEQ ID NO 46: AF262400. *Entamoeba histolytica* GI:11127699

5 RLHTVTGHDHNHSIQFDWSKCMGCGMCATKCTFGVLVKQPPKIPPFVQPNREKLSQ
ENTDKTRVLIDSECTGCGQCSLVCNFGSITPIDHLVDTFKAKEAGKKLVAMIAPSTR
LGVAEAMGMPIGSTAMAQLVHCLRLIGFDYVFDVDAGADKTTMDDYAEVIEMKKE
GKGPAITSCCPAWIELVEKEYPDLIPNVSTARSPIGCLAGCIKRGWAKDVGIAVEDLY
TVGIMPCIAKKTESQRQQIHQDYDASCTSNEIAAYFKKHLPPREECKFTQEREEALAKT
10 EDGQCCLPFRISGGSNIFGRTGGVCETVLRVIARNAGVDWNSCTVNKEETFKAAS
GSTMTNLSVDIGGTIITGAVCHGGYAIRHACELIRK GELKVDVEMMACVGGCLGG
AGQPKIPPAKKLEMDKRRVMLDILDQQTDIRAANENTDVLGWIDKHFDHQGAHQHL
HTYFTPRYQN

15

SEQ ID NO 47: AF242293. *Giardia intestinalis* GI:13506793

MPPKPQHDVTGVDSNNAIMIDYAKCIGCNMCIKACDVQGIGVYKQNEKPKYPPIVKL
20 STLFNSDCIGCGQCATICPVDAIAPKNNLEIYKGESASKKVRVALIAPSTRVAFGDVFG
LPIGTNTIYSLIRMLKQYLGFDYVFDVNF GADETTVIDTQELLHFKHEGRGPVFTSCC
PAWVNLCEMKYPELLPQVSTAKSCVAMVATLVKRRWVQEHLIPKGIVDSVDDVYV
ADIMPCTAKKDESMRPQLNRDVDICLTVREVAEHL YFLHGARLTLEEVEADALVLRP
GRSTQKKWDFDAPFNTVSGGSHIFGKTGGVAETCLRFISYMKKSPIENVKEELLKEFK
25 TPGQLVQTVKLVSCEIAGETYRALIAHGGSAINAAARMVLNKEVECDVVEQMACPG
GCQNGGGMPKIKGKKEAVLTRASTLDILDGKERFASAGENKTLWGFNGCLTEHEAH
ELLHTHYQHRPVESLLPQ

30

SEQ ID NO 48: NZ_AABB01000211. *Desulfitobacterium hafniense* GI:23112567

MVKIISITNNAKRQKGTSRKEKQAMKEVTKQQRIRVTVNGRQMEVYGDLTILQAL
LQEDIHIPHLCYDIRLERSNGNCGLCVVELGEGSEQQDVKACHTPIQEGMIIHTNSPRL

EHYRKIRLEQILADHNADCVAPCVMTCPANIDIQSYLSHAGNGNFETAIKVIKERNPF
PIVCGRVCPHSCEAQCRRNLIDEPVAINHVKRFIADWDIAHEQPWAPRKKAAATGKKI
AVVGAGSSGLSAAYYSAIQGHDTVFERHPRAGGMMRYGIPEYRLPKETLDREIGLI
ADLGVKIMTNKALGTHIRLEDLHQDFDAVYLAIGSWRATPLQIEGDNLEGVWLGINF
5 LEQVTKGADIKLGEHVVVIGGGNTAIDCARTALRKAGSVKL VYRRTREEMPAESY
EVEEAIHEGVEMYFLTAPHKIVAEGGRKLLHCIKMTLGEPRSGRRRPIPIEGSETAFE
ADTIIGAIGQSTNTQFLYHDLPVKLNKWGDIEINGKTMQTSEMNI FAGGDCVTGPAT
VIQAVAAGRHA AEAMDSFLMKG YVKEQPM DYSCSRGSLEDLPQWFEKIPRLKRAP
MPALPPAERRDNFRE VETGLSEETARAEARRCLKCGCYERYDCDLRQEASLHHVEF
10 KKPVHERPYIPIVEDHSIIIRDHNKCISCGRCIAACAEVEGPDILSFYMKHGRQLVGTK
SGLPLDQTD CVSCGQCVNACPCGALDYRSEIGRVFRAINDPGKTTVA FVAPAVRSVV
SSQYGVSYQEASRFIAGLLKKIGFDKVFDFTFAADLTIVEETTEFLTRLQSHKPIPQFTS
CCPGWVNFVERRYPEIIPYLSSCKSPQMMM GATVKITLRN

15

SEQ ID NO 49: AJ420003 *Nyctotherus velox* GI:19572306

ILFMEKNYPDMLNHLSTCKSPQGMLGALIKGYWAKNVKKIDPKDVVSVSIMPCTAK
20 KEEKDRITLKSDEGYNNVDYVLTTR ELAKMFKQSNIDPSKLPPTQFDNVMSEGTGAA
VIFGVT

SEQ ID NO 50: AC091811. *Oryza sativa* GI:18087682

25

MASSSSSASSRFSPALQASDLNDFIAPSQDCIISLNKGPSARRLPIKQKEIAVSTNPPEE
AVKJSLKDCLACSGCITS AETVMLEKQSLGDFITRINSDKAVIVSVSPQSRASLA AFFG
LSQSQVFRKLTALFKSMGVKAVYDTSSSRDLSLIEACSEFVTRYHQNLSSGKEAGK
NLPMLSSACPGWICYAEKTLGSFILPYISAVKSPQQAIGAAIKHHMVGKLGLKPHDV
30 YHVTVMPCYDKKLEAVRDDFVFSVEDKDVTEVDSVLTTEVL DLIQSRSVDFKTLE
ESPMDRLLTNVDDD GQLYGVSGGSGGYAETVFRHAAHVLFDRKIEGSVDFRILRNS
DFREVTLEVEGKPV LKFALCYGFRNLQNIIRKIKMGKCEYHFIEVMACPSGCLN GGG
QIKPAKGQS AKDLIQLLEDVYIQDVS VSNPFENPIAKRLYDEWLGQP GSENAKKYLH
TKYHPVVKSVASQLQNW

SEQ ID NO 51: AJ420004. *Psalteriomonas lanterna* GI:19572308

5 INLVEKHYPEYLPNLSSCRSPQGMLSSLIKNYWAKKMGIEPKDVVVVSFMPCGAKK
DEIKRPQLKGETDYVLTTRRELGKLFKMGGNLDLSVLEPVKYDDPLGESTGAAVIFGA
T

10 SEQ ID NO 52: AJ420001. *Nyctotherus ovalis* GI:19572302

IMFMEKNYPDMLNHLSTCKSPQGMLGALIKGYWAKNIKKMDPKDIVSVSIMPCTAK
KAEKERPQLRGDEGYKDVDYILTTRELAKMLKQSNIDLGKMEPTPFDKVMSEGTGA
AVIFGVT

15

SEQ ID NO 53: AJ420002. *Nyctotherus ovalis* GI:19572304

IMFMEKNYPDMLNHLSTCKSPQGMLGALIKGYWAKNVKKMDPKDIVSVSIMPCTA
20 KKAERKPQLRGDEGYKDVDYILTTRELAKMLKQSNIDLGKMEPRPFDKVMSEGTG
AAVIFGVT

25 SEQ ID NO 54: NZ_AAAG01000003 *Rhodospirillum rubrum* GI:22966612

MRPVQRPRRWPGRLRQLSPERPVDRRSRRRSGAARPGRRRGSGVQHEILRSVSQRD
MSMSIQPTVTIDPELCTGCGRCVETCPVQAIAGSRGKAHEIEAAACVSCGRCVATCA
AFDSIFDAFPTPRPVRLKRRGLPGSLKEPLFAAHDPSRIEAVRKAFATPKRMTVMQVD
30 TMACVALAEDFGLPPGSLPLKIASAARQLGFDRVYRTSFPAGLAVLETAHEMAARL
ANGGNLPVINSSCPAVVAFLERRYPELLHYLSTVKSPHQIAGALYNSYLADAANLAP
ANIHKVSVVACL SHKAEAEERPMMTCGCPDIDTVLTARELAILIKDAGIDVPLLGDGE
FDNDFPEIEGLDTLYCAPGDVSRAVLGAGRWFLGQGEGVGAPAGETVEVLDEATRL
TRLAYPGGTQLQALT VAGFDKAVPYLEAIKAGRNAFQFLEIASCPQGCASGAGLPKVL

LETEKPARYRARIENLPPAAPEAWSRLPGHPSIVALYGGYFGKAIGDKSNRRRLHTQY
AEPAAAP

5

SEQ ID NO 55: NZ_AABB01000211. Desulfitobacterium hafniense GI:23112569

MAVEKLTGEVLTDQLDYQEVRLQGIKEAAVEAKGKKVNVAVISGLHNVEPILEKII
EGMEVGYDLIEVMACPGGCICGAGHPVPEKIDTLEKRQQVLVNIDQTSRYRKSQENP
10 DILRLYDEYYGEANSPLAHKLLHTHYEAVKREPVAKHDRRMADSAFVTHELTLCCTC
DKCTAQGSRELFAALSGKIRKLKMDSFVTARTIRLKENHPGQGVYAAIDGKLIETPVE
QLEQRIFQHLIR

15 SEQ ID NO 56: NZ_AABB01000211 Desulfitobacterium hafniense GI:23112568

MVSIVPCIAKKYEAARPEFRSEGIRDVDAVLTSTEMLEMADIKLIEPADVEPQDFCEP
YKRVSGAGILFGASGGVAKRPCGWRWRN

20

SEQ ID NO 57: AY071338. Drosophila melanogaster GI:17945827

MSRLSRALQLTDIDDFITPSQICIKPVQIDKARSKTGAKIKIKGDGCFEESSEGNLKLNK
VDISLQDCLACSGCITSAAEVLITQQSREELLKVLQENSKNKASEDWDNVRTIVFTLA
25 TQPILSLAYRYQIGVEDAARHLNGYFRSLGADYVLSTKVADDIALLECRQEFVDYR
ENENLTM LSSSCPGWVCYAEKTHGNFLLPYVSTTRSPQQIMGVLVKQILADKMNP
ASRIYHVTVMPCYDKKLEASREDFFSKANNSRDVDCVITSVEVEQLLSEAQQPLSQY
DLLDLDWPWSNVRPEFMVWAHEKTLSSGGYAEHIFKYAAKHIFNEDLKTELEFKQLK
NRDFREILKQNGKTVLKFAIANGFRNIQNLVQKLKREKVSNYHFVEVMACPSGCIN
30 GGAQIRPTTGQHVRELTRKLEELYQNLPRSEPENSLTKHYNDFLDGFQSDKSYDVLH
TRYHDVVSELSISLNINW

SEQ ID NO 58: AL049559 *S. pombe* GI:4581520

MAKLSVNDLNDFLSPGAVCIKPAQVKKQESKNDIRIDGDAYYEVTKDTGETSELGIA
SISLNDCLACSGCITS AETVLVNLQSYQEVLKHLESRKSQEILYVSLSPQVRANLAAY
5 YGLSLQEIQAVLEMVFIGKLG FHAILDTNASREIVLQQCAQEFCNSWLQSR AHKNQN
QVTNSVVNEHPLPHSTS QISGVHSNTSSNSGINENAVLPILSSSSCPGWICYVEKTHSNL
IPNLSRVRSPQQACGRILKDWAVQQFSMQRNDVWHLSLMPCFDKKLEASRDEFSEN
GVRDVD SVLTPKELVEMFKFLRIDPIELTKNPIPFQQSTDAIPFWYPRITYEEQIGSSSG
GYMGYVLSYAAKMLFGIDDVGPYVSMNNKNGDLTEYTLRHPETNEQLISMATCYG
10 FRNIQNLVRRVHGNSSVRKGRVLLKKRVR SNAQNPTEEPSRYDYVEVMACPGGCIN
GGGQLPFPSVERIVSARDWMQQVEKLYYEPGTRSVDQSAVSYM LEQWVKDPTLTPK
FLHTSYRAVQTDNDNP LLLANKW

15

SEQ ID NO 59: AJ419999. *Metopus contortus* GI:19572298

IIFAEKNYPEMVNHLSTTKSPMQMLSSLSKGYWAKEGKKIDPKNVVNVAIMPCTAK
KAWKERPDMKADNGDPVTDYVLT TRELGTMLRQSNINPVSLPKTPFDKIMGESTGA
20 AVIFGAT

SEQ ID NO 60: AK076232 *Mus musculus* GI:26345230

25

MKCEHCTRKECSKSKNDDQENVSSDGAQPSDGASPAKESEEKGEFHKLADAKIFLS
DCLACDSCVTVEEGVQLSQQSAKDFLHVLNLNKRCDTSKHRVLVVSVCPSLPYFA
AKFNLSVTDASRRLCGFLKSLGVHYVFDTTIAADFSILESQKEFVRRYHQHSEEQREL
PMLTSACPGWVRYAERVLGRPIPYLCTAKSPQQVMGSLVKDYFARQQNLSPEKIFH
30 VVVAPCYDKKLEALREGLSTTLNGARGTDCVLTSGEIAQIMEQSDLSVKDIAVDTLF
GDMKEVAVQRHDGVSSDGHLAHVFRHAAKELFGEHVEEITYRALRNKDFHEVTLE
KNGEVLLRFAAAYGFRNIQNM IQKLKKGKLPYHFVEVLACPRGCLNGRGQAQTEDG
HTDRALLQQMEGIYS GIPVRPPESSTHVQELYQE WLEGTESPKVQEV LHTSYQSLEPC
TDGLDIKW

SEQ ID NO 61: NM_065691 *Caenorhabditis elegans* GI:25148541

5 MEDSGFSGVVRLSNVSDFIAPNLDCIIPLETRTVEKKKEESQVNIRTKPKDKESSKTE
EKKS VKISLADCLACSGCITSAETVLVEEQSFGRVYEGIQNSKLSVVTVSPQAITSIAV
KIGKSTNEVAKIIASFRRRLGVKYVIDSSFARKFAHSLIYEELSTTPSTSRPLLSSACPGF
VCYAEKSHGELLIPKISKIRSPQAISGAIIKGFLAKREGLSPCDVFHAAVMPCFDKLE
ASREQFKVDGTDVRETDCVISTAELLEIIKLENDEAGDVENRSEEEQWLSALS KGSV
10 IGDDGGASGGYADRIVRDFVLENGGGIVKTSKLNKNMFSTTVESEAGEILLRVAKVY
GFRNVQNLVRKMKTKKEKTDYVEIMACPGGCANGGGQIRYETMDEREKLIKVEA
LYEDLPRQDDEETWIKVREEWEKLDKNYRNLLFTDYRPVETNVAQVLKW

15 SEQ ID NO 62: AK081305 *Mus musculus* GI:26349103

MKCEHCTRKECSKSKTDDQENVSSDGAQPSDGASPAKESEEKGEFHKLADAKIFLS
DCLACDSCVTVEEGVQLSQQSAKDFLHVLNLNKRCDTSKHRVLVVSVCPSL PYFA
AKFNLSVTDASRRLCGFLKSLGVHYVFDTTIAADFSILESQKEFVRRYHQHSEEQREL
20 PMLTSACPGWVRYAERVLGRPIPYLCTAKSPQQVMGSLVKDYFARQQNLSPEKIFH
VVVAPCYDKKLEALREGLSTTLNGARGTDCVLTSGEIAQIMEQSDLSVKDIAVDTLF
GDMKEVAVQRHDGVSSDGH LAHVFRHAAKELFGEHVEEITYRALRNKDFHEVTLE
KNGEVLLRFAAAAYGFRNIQNM IQKLKKGKLPYHFVEVLACPRGCLNGRGQAQTEDG
HTDRALLQQMEGIYS GIPVRPPESSTHVQELYQEWLEGTESPKVQEV LHTSYQSLEPC
25 TDGLDIKW

SEQ ID NO 63: AJ419998. *Neocallimastix* sp GI:21425619

30 IMFAEKNFPDMVNNLSTTKSPMQMLSSLTKGYWAKDIKKINPKDVVNVAIMPCTAK
KQEKDRPGMKTDEGDKVTD FVLTTRELGMMLRQANIDPTKLPGTKFDKVMGESTG
AAVIFGAT

SEQ ID NO 64: AJ420000. *Nyctotherus ovalis* GI:19572300

5 IIFMEKNYPDMLSHLSTCKSPQGMLGALIKGYWAKKVKKVDPKDVSVSIMPCTAK
KAEKERPQLRGDEGFKDVDYVLTTRRELAKMLKQSNIDLGKVEPTPFDAVMSEGGA
AVIFGVT

10 SEQ ID NO 65: NC_003366. *Clostridium perfringens* GI:18311557

MAIKDANKQYIKFDTAVQVLKYEVLKRIAEKEFDGTLDKEKLNIAKEIVDDLKPNVR
CCIKERAIVEERMKLALGGHENRENMIIEVIDIACDECPVNRFIVTDACRGCLAKKCR
DSCNFGAISFDNRKCKIDYEKCKEKGKCKEVCYPYNAIAEVKRPCMRACIPKALSVDV
15 DSKKAVIDDSKCIQCGACVVDPCFGAIMDKSYLVDVIRLLKDEKKV
YAIVAPAISSQFNHSGKIGKVITAIKKLGFEDVFEAALGADLVAVHECNEFKEKGELDF
MTTSCCPAFVSYIEKNYPELKECISNTVSPMVAMARLIKSONKDVKTFIGPCIAKKT
EAKRNEVSGDVDYVLTFEELLALLDSRNIKIDECEESDTKHGSFYGRLFARSGGVTE
VKHLIDSEGIKVDFRPILGDGIKDCDIKLRLAKLKRAQGNFLEGMACKGGCINGPGSL
20 NHDIKNSKEVDKYGELSSSEKIKDTLADIKFEDLNLSKNE

SEQ ID NO 66: NM_012336. *Homo sapiens* GI:6912524

25 MKCEHCTRKECSKKTCTDDQENVSAADAPSPAQENGEKGEFHKLADAKIFLSDCLAC
DSCMTAEEGVQLSQQNAKDFFRVLNLNKKCDTSKHKVLVSVCPQSLPYFAAKFNL
SVTDASRRLCGFLKSLGVHYVFDTTIAADFSILESQKEFVRRYRQHSEEERTLPMLTS
ACPGWVRYAERVLGRPITAHLCCTAKSPQQVMGSLVKDYFARQQNLSPEKIFHVIVAP
30 CYDKKLEALQESLPPALHGSRGADCVLTSGEIAQIMEQGDLSVRDAAVDTLFGDLKE
DKVTRHDGASSDGH LAHIFRHA AKELFNEDVEEVTYRALRNKDFQEVTLKNGEVV
LRF AAAYGFRNIQNMILKLKKGKFPFHFVEVLACAGGCLNGRGQAQTPDGHADKAL
LRQMEGIYADIPVRRPESSAHVQELYQEWLEGINS PKAREVLHTTYQSQERGTHSLDI
KW

SEQ ID NO 67: BC016440. Homo sapiens GI:16741189

5

MKCEHCTRKECSKKTKTDDQENVVSADAPSPAQENGEKCDTSKHKVLVVSVCPSLP
YFAAKFNLSVTDASRRLCGFLKSLGVHYVFDTTIAADFSILESQKEFVRRYRQHSEEE
RTLPM LTSACPGWVRYAERVLGRPITAHLC TAKSPQQVMGSLVKDYFARQQNLSPE
KIFHVIVAPCYDKKLEALQESLPPALHGSRGADCVLTSGEIAQIMEQGDLSVRDAAV
10 DTLFGDLKEDKVTRHDGASSDGH LAHIFRHA AKELFNEDVEEV TYRALRNKDFQEV
TLEKNGEVVLRFAAA YGFRNIQNMILK LKKGKFPFHFVEVLACAGGCLNGRGQAQT
PDGHADKALLRQMEGIYADIPVRRPESSAHVQELYQEWLEGINS PKAREVLHTTYQS
QERGTHSLDIKW

15

SEQ ID NO 68: NM_031968. Homo sapiens GI:14165461

MKCEHCTRKECSKKTKTDDQENVVSADAPSPAQENGEKGEFHKLADAKIFLSDCLAC
20 DSCMTAE EGVQLSQQNAKDFFRVLNLNKKCDTSKHKVLVVSVCPSLPYFAAKFNLS
SVTDASRRLCGFLKSLGVHYVFDTTIAADFSILESQKEFVRRYRQHSEEERTLPM LTS
ACPGWVRYAERVLGRPITAHLC TAKSPQQVMGSLVKDYFARQQNLSPEKIFHVIVAP
CYDKKLEALQESLPPALHGSRGADCVLTS EISQAWWCTPVITATREAAARESLEPGR
QRLQRDKIAPLDSSLGGGGEIAQIMEQGDLSVRDAAVDTLFGDLKEDKVTRHDGASS
25 DGH LAHIFRHA AKELFNEDVEEV TYRALRNKDFQEV TLEKNGEVVLRFAAA YGFRN
IQNMILK LKKGKFPFHFVEVLACAGGCLNGRGQAQTPDGHADKALLRQMEGIYADI
PVRRPESSAHVQELYQEWLEGINS PKAREVLHTTYQS QERGTHSLDIKW

30 SEQ ID NO 69: AE015936. Clostridium tetani GI:28202388

MHNDYREIFKRLSKSYDDTFEKEVENILSSHSM DREKLAKIISILCGVNIEHSENYIS
NLKNAIKNYTASAEKVVTKLPCSTQCAKDGDII CEKSCPVNAIFRDPNDNNIYINDEL
CLDCGLCVRNCPSGSILDKKEFIPLAELLKSESIVIAAVAPAIMGQFGENTTINQLRTAF

KKLGFTDMVEVAFFADMLTLKEAVEYDHFVKDEQDFMITSCCCPMWVGMLKKVY
NDLVKYVSPSVSPMIAAGRVLKLLNPNCKVVFVGPCLAKKAEAREKDLLGDIDFVLT
FTELRFDFVFDIQPENLEEDFSSEYASKGGRLYARTGGVSIIVSEAIEKLFPNKKYKFL
KTIQADGVKGCKSLDKIKQEDISANFVEGMGCVGGCVGGPKVIIDPSEGRNAVNNF
5 AENSSIKVSVDSNCMNDILSKININSVEDFKDKDKISIFEREFK

SEQ ID NO 70: NZ_AABI01000028. *Desulfovibrio desulfuricans* GI:23475858

10 MYFRTYDNTINFEIMVRIAKAFHGDSFEEQVARIPLEMRRPKAHSSRCCIYRDRAIRY
RCMAMLGYAIEDETDELTSLSQYAKGALERDSIQGSMLTFIDEACNGCVRTHYEATS
ACRGCLAEACVQHCPKDAVRIVDGKSRIDPDKCVQCGKCMNVCPYHAIVQIIPCEE
SCPTGAISKDECGKQVIDYDRCIFCGKCMAACPFAAVLEKSQMIDVLRRIREGRKVV
15 AIVAPAIAGQVQAPMSRLATALRQLGFADVAEVASGADTTARLEADEFVERMEHGA
AFMTSSCCPAYTQLVDKHLPELAPFVSDTRTPMHYTAAMVKDHDPMVTVFIGPCV
AKRNEGKHDELVDHVLTFQEMVAMLTAAGISVDACEDGRFMFPAMREGRSFPVSG
GVTAGVQAHIGTRAIEVRPLSVDGLNKKTFRQLKTWAKKGCEGNFVEVMGCQGGCV
AGPAIVM

20

SEQ ID NO 71: AE015944. *Clostridium tetani* GI:28204446

25 MIVFENQLKKLYLVLKEVAKMTLEDRLGEEDIERISFDIIGDKAEYRCCVYKERAI
VYERAKLATGCLPNGQVAEEFVHVEDDDQIYVIDAACDKCPINKYVVTEACRGCLQ
HKCMEVCPAGSINRAAGKAYINHETCKEGLCESACPYNAIAEVMRPPRRACPTGA
LQMNLEDNKAINKEDCINCGSCMSVCPFGAISDKSYIVDITKALKNNKKVYAMVAP
AITGQFGKDVSVGKMKNAFKAMGFEDMLEVACGADAVAAHESEEFIERLESKKY
MTTSCCPGFLGYIEKKFPDQLENVSNTVSPMVAIGRMIKKEYEDSVVVFVGPCTAKK
30 AEIKRKGIKDAVDYVMTFEEIAALMGAFEIDPAECEEEEDINDGSNYGRGFAQGGGVV
SAIQNCIKDKEGIKFNPLRVSGPDQIKRAMIMAKVGKLSNFIEGMMCEGGCIGGPAT
MVSAVKAKAPLMKFSKSSTIKDVKDNEVLDDKYKDINMER

SEQ ID NO 72: AK117970. *Arabidopsis thaliana* GI:26451001

MDLIKLGVDKFDLEESPLDRVLTNVTEEGDLYGVAGSSGGYAETIFRHAALKALFGQ
5 TIEGPLEFKTLRNSDFREVTLQLEGKTVLKFALCYGFQNLQNIVRRVKTRKCDYQYV
EIMACPAGCLNGGGQIKPKTGQSQKELIHSLEATYMNDDTLNTDPYQNPTAKRLFEE
WLKEPGSNEAKKYLHTQYHPVVKSVTSQLNNW

10 SEQ ID NO 73: NC_003366. *Clostridium perfringens* GI:18309258

MNKKYNSLKFELISSYYSEDNFDEKLNDIVKNNFNSKEDAIEVLSSLCGVDIDKNSDN
IAYDIRKAITTHKIKKNIVDKVSVCTKNCSKESKGKCQSLCPFDAILTDPIDNSKYIDP
15 NLCQNCGICVQVCESGHFLDRIELLPIIDLIKNNETVIAAVAPAIAGQFGENVSLDMLR
EAFIKIGFSDMIEVAFADMLSIKEAVEFNHHVEKTGDILITSCCCPMWVAMLRKCY
KDLVKDVSPSVSPMIAAGRVIKKLNKDAKVVFIPGCIKKAAREKDLVGAIDYVLT
FEELNGIFEALKIDPSSMKGVPSIEYTSRGGRLYARTGGVSEAINDVVKELYPDKAKIF
KAVQANGVKECKELLNKVQSGELKANFIEGMGCVGGCVGGPKRIVDPSIGKKHVDE
20 VAYNSPIKVATHSHTMDEVLLRLGINSLSKFEDKEKISIFEREF

SEQ ID NO 74: NZ_AABB01002911. *Desulfitobacterium hafniense* GI:23120640

25 MAQSEIMKIRRQVLKSALDWVSHDQNRKDRATLARQIIPDGTPRYRCCIHKERA
ERLKAVLEPDEGPVIRVLKEGCNGCEMHRYSVTDHCQNCVGHFCFTNCPKKAILFIN
NKAFIGDQTRCVECGLCARNCPYHAIIEYRRPCEDSCPTKAISVREDRIASIAEAHCTSC
GKCIISCPFGAVAESSQLIHLFEAVRNPEHKIYAVIAPAFVGQFGRKVSPGQVKSALLK
30 LGFQDVLEAALGADRTIELEAREYDERLAHGEEFMTSSCCPAYVSAVIKEKPDLFHHI
SSTLSPMAQVAHILKEKDPEAKIAFIGPCVAKKEEGKRPETKVDFVLTFEELMVWLD
YAGINPAEESEQ

SEQ ID NO 75: NZ_AAAS01000012 *Geobacter metallireducens*...gi:23055369

MCHWLHREAGLVYDPAVDQAINRVSGLTLSAGRTMEPIITVKEKCRKCYCCVRSCP
VKAIVAKSYTEIIVDRCIGCGNCLSNCPQQAQKMOVADKVGVTKEKLLSSGEEVIAVLG
5 SSFPAFFHNVTPGQLVAGLRKIGFAEVHEGSYGAELIADDYARITSEKGHPRISSHCPA
IVDLIERHYPKLVGNLVPVVSMPVAMGRYLKGTLLGQHVRVVYISSCVANKLETQTQ
ETRGAVDIVLTYRELEGIFRSRQIALPALADEPLDGIRPGAGRLFPIA
DGTFRAGFIPFDPLDTEIVAACGEVNVMMGIINDLAAGRISPRIADLRFCYDGCIGGPGR
NRALTEFYRRNRVIAHFKQEVPCRTVPNSLLEAGRVSFGRSFASKYAKLEAPKANDV
10 RKILNATNKFTVKDELNCRACGYRTCNEYAVAVFQGLAEIEMCLPYNLQQLEEDRG
RLIQKYELARRELEREYGDEFIVGNDRKTLDELGLIKQVGPTPTTVLIRGESGTGKEL
TARAIHRYSKRNDKPLVTVNCTTTITDSLLESELFHKGKRGFTGAVADKKGLFEAADG
GTIFLDEIGDITPKLQAE LLRVLD MGEVRPVGGTAAKKVDVRLIAATNKNLEQGVRE
GWFREDLYYRLNVFTITMPPLRSRVESIPILVHHFMDKASTKLNKRMVGIEDRAVKA
15 LTKYPWPNGNIREMQNVIERAAVLTHDGVIRVENFPLALSEGLEEGFATGLDIIHAASFR
SREQHMGKLEKKLIQRYLTEANGNISRAAKLANIPRRTFYRLLDKYRLRERDVR

20

SEQ ID NO 76: AE007819 *Clostridium acetobutylicum* GI:15026303

MNNKYIELFKSLVDSYYNDTFDSFVYHILSDEEVDKKELSKVISSLCGVSVEFKDTET
YISELKKAISNYKCTDNIVEKIKECDSSCHSNEGETPCQKSCPFDAILVDKNTKTSHIQ
25 KDLCTDCGNCTSCPSGSILDKIEFMPLLNLFKNNETVIAAVAPAIAGQFGENVSLEML
RTAFKKVGFADMVEVAFFADMLTIKEAFEFNELVNSKDDLMTSCCCPMWVSMIRKI
YKDLARHVSPSVSPMIASGRVIKKLNPNCVVFIGPCIAKKAESRSQDISDAIDFVLTF
EELKGIFDVLDIDPEKLPETHTKSYASREGRLYGRTGGVSTS
VDEAVKRIFPNKHHLFKSTKVDGVKDCKDILNKTQAGNIGANFLEGMGCVGGCVGG
30 PKAIVHKDQGRESVNKTAESSEIKISVDSERMKDILSRIGINSIEDFGDKSKVDIFERRF

SEQ ID NO 77: NC_004347 *Shewanella oneidensis* GI:24375409

MNKKKHLFAEDSFFLSRRKFMAVGAAFVAALAIPIGWFTSKLERRNEYIKARSQGLY
KDDSLAKTRVSHANPAVEKYYKEFGGEPLGHMSHELLHTHFVDRTKLSS

5

SEQ ID NO 78: AY172963. *Entamoeba histolytica* GI:27652439

MSTQLTPLRNKIISEVVKCFKSGRFIEDIDKLPTILTDGDGWKPTSKFVHSREQEEGIY
REKVL SVLGFDGEYDDITPLHVYAQKALERTSLHEPVFGISQKGCNKCHFNGYFVT
10 QACEGCTSRPCSVNCPKKCISFGEDGRAVINQNNCIKCGRCYKFCPYGAISKSVPCV
KACPCGAMLDSPGKVKTIDFEKCINCGGCMRACPF GAILPRSNLIDVLKILPTKKVVA
CPAPSIAAHFGKYDLALVSGGLIQVGFTSVEDVSYGADLCALNEAKEFEERIVKNKK
DFMTTSCCPAYINAINKHMPELKENVSHTPTPMHFATQAVKDRDQETVTVFIGPCNA
KRWETLQDSTTDYCLTFDEIFGLFEGSGIDLSKVQPYTFVDKAHKEGKIFAVSGGVAS
15 AVASLLPKEVPDGVIKPTIIDGFSQENFKRLKNFKKNITGNLVEVMVCEGGCAYGPG
CPGLNTPATSAKIKIAVDKMEAHPEGRWVGLPNSQIKPIKVEN

20 SEQ ID NO 79: AY145118. *Cryptosporidium parvum* GI:23213133

MFSTAVKLANLDDYLESSQDCIVSLLSDKDDTKPKIAVMRPAKAQGNKDDKKSGTS
DKATVNVADCLACSGCVTSAEAKLLEDQNVSEFMNILKQKRLTVVSISNQSCSSFAC
HLNCDLITIQRKLSGLFKHIGARFVMNSTISEYISLLETKYEFISRYKAKSDLPMIISHCP
25 GWICYSEKSLNSSVLPLLSKVRSAQQLQGILIKTLTLEIYNQLLFLYKFRLSNSYRTNM
NVKSTFTQNDNFVEQSDIFHVAIMPCHDKKLESTRSSLSLKSS
DKNSSCPEVDIVLATSEVGEIIKLAGFNSLLDVPEAPLDNLWLNQNFQITKHNLSLLI
TENYVSNQILNQFSWLIPSYFNNSNGGFCEYIIRSAIKELAGDHIDNKVQLPFNKLKND
ILEAKYIKNNVELNYCLAYGFRAIQSISRKLNQKNASQNTQYKQSVVNHVNYHLIE
30 AMACPTGCVSGGGQILSQNDQNDNDNSDLNKLKRNKIFIDEVQEALYKGINLNKNQE
VILPDEIPIVNILYEYLIHIDKQIDRSSGLKLPFLRNDFVSINEVPTASSLKW

SEQ ID NO 80: X70373. *Kluyveromyces lactis* GI:5538

MSALLRDADLNDFISPLACVKPAQPKVEKKPSFEVEVGIESSEPEKVSISLQDCLA
CAGCITSSEEILLSKQSHKVFLEKWSELEELDERSLAVSISPQCRLSLADYYSMCLADL
DRCFQNFMTKFNKAKYVVGTFGRSISISRINATLKDRVPENEGPLLCVCPGFVLYA
5 EKTKEPILPHMLDVKSPQQITGNLLKQADPTCYHLSIMPCFDKKLEASREECEKEVDC
VITPKQFVAMLGDLSDFKSYMTEYDSSKELCPSGWDYKLHWLSNEGSSSGGYAYQ
YLLSLQSSNPESDIITIEGKNSDVTEYRLVSKSKGVIASSSEVYGFRNIQNLVRKLSQSA
SVKKRGIKVKRRGQSVLKSGETSEKTTKVLTADPAKTDVFEVMACPSGCINGGGLLN
EEKNANRRKQLAQDLSLAYTKVHSVNIPIVHAYDDKSNDFKYNLRVIEPSTSSDVV
10 AVGNTW

SEQ ID NO 81: NC_003232. *Encephalitozoon cuniculi* gi:17158048
15 MDALIRPPMSFFADLPKDNKKCIKIGSPLALSLSLCLACSGCVSADEAGALSEDLSFV
LDLSPQTSFVLSPQSKINIFNLYREDGMEYREFEAVLSSFLRSKFNIHRIVDTSYLRSKI
YEETYREYMATNHLIVSACPGVVITYIERTAPYLIGYLSRVKSPQQMAFSLVKGSRTV
SVMPCQDKKLENGRDGVKFDFILTTRGFCKALDSLGFRRPARASGKSLCSMEEAETT
20 QWNIGTSSGGYAEFILGKHCVVETREIRNGIKEHLLDDGRTISQITGLENSINYFKSSK
TKGPRHKMTEIFLCKNGCIGGPGQERVNDVEMDIREYDRNGREQPRIFYSSPGLEEKR
VFREVKAKRVDLRVDW

25 SEQ ID NO 82: AF312932 *Tritrichomonas foetus* GI:14164621
MCIKACNSVAGQGVLKLVKVGNNKKLVSTKSGKPLQETNCIKCGQCTLVCGPGALTQ
KDAIQTVSEVLKNPGDKVLVCQTAPAIRXNLADGLGMPAGSIITGKMVTALKMLGF
KYVFDTNFGTDXTIG

30

SEQ ID NO 83: AJ271546 *Scenedesmus obliquus* GI:13311187

MPEWQPGGRYAVSVRPPVNRRRAVVAAERRRLVVRAAGPTAECDCPPAPAPKAPHW
QQTLDLAKPKEQRKVMIAQIAPAVRVAIAETMGLNPGDVTVGQMTGLRMLGFD
YVFDTLFGADLTIMEEGTELRLHRLQDHLEQHPNKEEPLPMFTSCCPGWVAMVEKSN
PELIPYLSSCKSPQMMLGAVIKNYFAAEAGAKPEDICNVSVMPVCVRKQGEADREWF
5 NTTGAGGANVDHVMTTAELGKIFVERGIKLNLDQESPFDPVGEESGGGVLFGTG
GVMEAALRTVYEVVTQKPLDRIVFEDVRGLEGIKESTLHLTPGTPSPFKAFAGADGT
GITLNIAVANGLGNAKKLIKQLAAGESKYDFIEVMACPGGCIGGGGQPRADKQILQ
KRQAAMYDLDERAVIRRSHENPLIGALYEKFLGEPNGHKAHELLHTHYVAGGVPDE
K

10

SEQ ID NO 84: AAAB01008984 *Anopheles gambiae* GI:21302659

SRFSSALQLTDLDDFITPSQECIKPVKIETSKSKTGAKITIQEDGSYVQESSSGIQKLEK
15 VEITLADCLACSGCITSAEGVLISQQSQEELLRVMNANNLAKLNNQRDEIKFVFTVS
QQPILSLARKYNLTPEDTFEHIAGYFKKL GADMVVDTKIADDLALIECRNEFIERYNT
NRKLLPMLASSCPGWVCYAEKTHGNFILPYIATTRSPQQIMGVLVKQYLAKQLQTTG
DRIYHVTVMPCYDKKLEASREDFSEVENS RDVDCVITSIEIEQMLNSLDLPSLQLVE
RCAIDWPWPVTRPSAFVWGHESGSGGYAEYIFKYAARKLFNVQLDTVAFKPLRNN
20 DMREAVLEQNGQVLMRFAIANGFRNIQNMVQKLKRGKSTYDYVEIMACPSGCLNG
GAQIRPEEGRAARELTAELECMYRSLPQSTPENDCVQTMATFFDSEGLNKRQSL
HTSYHQIEKINSALNIKW

25 SEQ ID NO 85: NC_004347 *Shewanella oneidensis* GI:24375408

MTTTYQPGEIQGLIKINASKCKGCDACKQFCPTHAIN GASGAVHSIDEDKCLSCGQC
LINCPSAIEETHSALETVIKKLADKNTTVVGILAPAVRVAIGEEFGLGTGELVTGKLY
GAMNQAGFKIFDCNFAADLTIMEEGSEFIHRLHANVKGEANAGPLPQFTSCCPGWVR
30 YLETRYPALLPNLSTAKSPQQMAGTVAKTYGAKVYQMOPENIFTVSVMPTSKKLE
ASRPEFNSAWQYHQEHGANS PSYQDIDAVLTTREMAQLLKLLDIDLANTA EYQGDS
LFSEYTGAGTIFGTTGGVMEAALRTAHKVLTGTEMAKLEFEPVRGLKGVKSASVSLF
DTELNQDVTNVAVVHDMGN NIEPVLRDVMAGTSPYHFIEVMNCAGGC VNNGGQ
IEGKGSSWLGN

SEQ ID NO 86: NZ_AABG02000026 *Clostridium thermocellum* GI:23021197

5

MAFVWRNVRSRPFPPKPNGRGCEKMQMNVNTIDNCKIQVPANYTVLEAAKQANIDI
PTLCFLKDINEVGACRMCVVEVKGARSLQAACVYPVSEGLEVYTQTPAVREARKVT
LELILSNHEKKCLTCVRSENCCELQRLAKDLNVKDIRFEGEMSNLPIDDLSPSVVRDPN
KCVLCRRRCVSMCKNVQTVGAIDVTERGFRTTVSTAFNKPLSEVPCVNCGQCINVCVP
10 GALREKDDIDKVWEALANPELHVVVQTAPAVRVALGEEFGMPIGSRVTGKMVAAL
SRLGFKKVFDTDTAADLTIMEEGTELINRIKNGGKLPLITSCSPGWIKFCEHNYPEFLD
NLSSCKSPHEMFGAVLKSYYAQKNGIDPSKV FVVSIMPCTAKKFEAQRPELSSTGYP
DVDVVLTTRELARMIKETGIDFNSLPDKQFDDPMGEASGAGVIFGATGGVMEAAIRT
VGELLSGKPADKIEYTEVRGLDGIKEASIELDGFTLKA AVAHGLGNARKLLDKIKAG
15 EADYHFIEIMACPGGCINGGGQPIQPSSVRNWKDIRCERAKAIYEEDESLPIRKSHENP
KIKMLYEEFFGEPGSHKAHELLHTHYEKRENYPVK

20 SEQ ID NO 87: NZ_AABB01000275 *Desulfitobacterium hafniense* GI:23113285

MTMGQLRAALKHLGFYGMIEVALFADVLSLKEALEFDKHHVQTDKDFVLTSCCCPIW
VGMVKRVYDTLVPHISPSVSPMVACGRGIKRLHPDAKTVFIGPCIAKKAEAKEPDIRD
AVDAVLTFHELKQIFEATDIEPSEMEDIPSEHSSTSGRIYARTGGVSKSISDTLNRIRPD
25 KPVKIKSIQANGIKECKALLNDIMNNEIKANFYEGMGCPGGCVGGPKAIVDVDRGTE
FVNKYGAEADALTPADNQHVLELLKQLGIDSVEELLGGESAAIFQRDF

SEQ ID NO 88: AAR04931. *C. reinhardtii* iron-hydrogenase gi:37911261

30 MALGLRAELRAGQAVACARRTNAPAHPAAVVPVLPSRGDKFFNLSQKVPSSQPARG
STIRVAATATDAVPHWKLAL EELDKPKDGGRKVLI AQVAPAVRVAIAESFGLAPGA
VSPGKLAAGLRALGFDQVFDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMF
TSCCPGWVAMMEKSYPELIPFVSSCKSPQMMMGMAMVKTYLSEKQGIPAKDIVMVSV
MPCVRKQGEADREWFCVSEPGVRDVDHVITTAELGNIFKERGILPELPDSDWDQPL

GLGSGAGVLF GTTGGVMEAAVRTAYEIVTKEPLPRLNLSEVRGLDGIKEASVTLVPA
PGSKFAELVAARLAHKVEEAAAAEAAA AVEGAVKPPIAYDGGQGFSTDDGKGGLK
LRVAVANGLGNAKKLIGKMVSGEAKYDFVEIMACPAGCVGGGGQPRSTDQITQK
RQAALYDLDERNTLRRSHENEAVNQLYKEFLGEPLSHRAHELLHTHYVPGGAEADA

5

SEQ ID NO 89: gi:37911254 AAR04930 *C. reinhardtii* iron-hydrogenase HydA2

10 MALGLRAELRAGQAVACARRTNAPAHAAVVPVLPSRGDKFFNLSQKVPSSQPARG
STIRVAATATDAVPHWKLAL EELDKPKDGGRKVLIAQVAPAVRVAIAESFGLAPGA
VSPGKLAAGLRALGFDQVFDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMF
TSCCPGWVAMMEKSYPELIPFVSSCKSPQMMMAMVKTYLSEKQGIPAKDIVMVS
MPCVRKQGVADREWFVSEPGVRDVDHVITTAELGNIFKERGIILPELPDSDWDQPL
15 GLGSGAGVLF GTTGGVMEAAVRTAYEIVTKEPLPRLNLSEVRGLDGIKEASVTLVPA
PGSKFAELVAARLAHKVEEAAAAEAAA AVEGAVKPPIAYDGGQGFSTDDGKGGLK
LRVAVANGLGNAKKLIGKMVSGEAKYDFVEIMACPAGCVGGGGQPRSTDQITQK
RQAALYDLDERNTLRRSHENEAVNQLYKEFLGEPLSHRAHELLHTHYVPGGAEADA

20 SEQ ID NO 90: gi:15642974 *T. maritima*

MRRFFKNNLRNLSQNGETNSVRRCFALADVTVVINGRTLTPDNLTVIEACEKAGIEI
PALCHHPRLGESIGACRVCVVEVEGARNLQACVTKVRDGMVIKTSSDRVKTARKF
NLALLLSEHPNDCMTCEANGRCEFQDLIYKYDVEPIFGYGTKEGLVDRSSPAIVRDLS
25 KCIKCQRCVRACSELQGMHIYSMVERGHRITYPTGTPFDMFVYETDCIGCGQCAAF CPT
GAIVENSAVKVVLEELEKKEKILVVQTAPSVRVAIGEEFGYAPGTISTGQMVAALRR
LGFDYVFDTNFGADLTIMEEGSEFLERLEKGDLEDLPMFTSCCPGWVNLVEKVYPEL
RTRLSSAKSPQGMLSAMVKTYFAEKLGVKPEDIFHVSIMPCTAKKDEALRKQLMVN
GVPAVDVVLTTRELGKLIRMKKIPFANLPEEEYDAPLGISTGAAALFGVTGGVMEAA
30 LRTAYELKTGKALPKIVFEEVRGLKGVREAEIDL DGGKKIRIAVVHGTANVRNLVEKIL
RREVKYHFVEVMACPGGCIGGGGQPYSRDPEILRKRAEAIYTIDERM TLRKSHENPAI
KKLYEEYLEHPLSHKAHELLHTYYEDRSRKKRLAVK

SEQ ID NO 91: *C. reinhardtii* gi:9837540

MSALVLKPCAAVSIRGSSCRARQVAPRAPLAASTVRVALATLEAPARRLGNVACAA
AAPAAEAPLSHVQQALAEAKPKDDPTRKHVCVQVAPAVRVAIAETLGLAPGATTP
5 KQLAEGLRRLGFDEVFDTLFGADLTIMEEGSELLHRLTEHLEAHPHSDEPLPMFTSCC
PGWIAMLEKSYPDLPYVSSCKSPQMMLAAMVKSYLEKKGIAPKDMVMVSIMPCT
RKQSEADRDWFCVDADPTLRQLDHVITTVELGNIFKERGINLAELPEGEWDNPMGV
GSGAGVLFGTTGGVMEALRTAYELFTGTPLPRLSLSEVRGMDGIKETNITMVPAPG
SKFEELLKHRAAARAEAAAHGTPGPLAWDGGAGFTSEDGRGGITLRVAVANGLGN
10 AKKLITKMQAGEAKYDFVEIMACPAGCVGGGGQPRSTDKAITQKRQAALYNLDEKS
TLRRSHENPSIRELYDTYLGEPLGHKAHELLHTHYVAGGVEEKDEKK

15 SEQ ID NO 92: *T. tencongensis* NP_622546.1

MDKVRVTIDGITVEVPSYYTVLEAAKEAGIDIPTLCYLKEINQIGACRICLVEIEGVRN
LQTSCTYPVFDGMKVYTNTPKIREARRLNLELILSNHNRNCLTCVRSTNCELQALAK
RLGVEEIRFEGENIKYPIDDASPAVVRDPNKCVCRCVAVCSEVQNVFAIGMVNRG
20 FKTMOVAPSFGRSLKDSPCISCGQCIMVCPVGAIYEKDHTKRVYEALADDKKYVVAQ
TAPAVRVALGEEFGMPVGTIVTGKMAAALRRMGFDAVFDTNFAADLTIMEEGSELL
ERIKHGGKLPMITSCSPGWIAFCEKYYPEFIDNLSTCKSPHMMMGALVKSYYAEKKG
LDPKDIFVVSIMPCTAKKLEIEREEMIRNGMKDVEDAVLTRELARMIKEMGIDFVNL
KDEEFDEPLGMSTGAGAIFGATGGVMEALRTVAEIVEGRDIGKIDFEEVRGLEGVR
25 EATITIDGMDIKIAIANGTGNACKLLDKVKAGEVEYHFIEVMGCPGGCIMGGGQPIH
NPNEMEEVKKLRAKAIYEIDKNLPIRKSHENPAIKRLYEEFLGYPLSEKSHELLHTHY
SRKELYPLVK

30 SEQ ID NO 93: *N. frontalis* AAK60409.1

MSMLSSVLNKAVVNPKLTRSLATAAAEKMVNISINGRKFQVKPKTTVLEAAKANGY
YIPTLCYHQELPVAGNCRLCLVYAKGSWKPLTACTTEVWEGMEIETDSPAVIETVRS
SLSMMREEHPNDCMTCGSNGDCEFDLIYRYQIDAKHPVRSLLKHKSCKTNHSITEP

CYSPFDNTTFSVARDMNKCVKCGRCIRACHHFQNNILGFINRAGYERVGTPMDRPM
NFTKCVCEGQCSQVCPVGAITARTEVVVDVLRHLDTKRKVVCSTAPAIRVAPAEFEFS
TEADFDFTGKMVAGLRKLGFDYIFDTNFSADLTIMEEGTELIDRLNNGGKFPMFTSC
CPGWINMVEKSYPELSDNLSSCKSPQQMIGAVIKSYFAKKLGLSTEDIIHVSIMPCTA
5 KKGEARRPEFVQKGKDGDYPDIDYVITTRELLTLLKLLKKNPAELPDDKFDSPLGIG
SSAGNLFVGTGGVMEAAIRTAQVITGVENPIPLGELKAIRGLDGIIKAAANVPLKTKDG
KEVSVRAAVVSGGANIQKFLEKIKNKELEFDFIEMMMCPGGCINGGGQPKSADPEIV
AKKMQRMYTMDQAKLRLCHENPEIIDVYKNFLGEPNSHLAHELLHTHYNDRSKTI
HDMGHHEKK

10

SEQ ID NO 94: AAD33071.1 C. thermocellum gi:4927278

MVNVTIDNCKIQVPANYTVLEAAKQANIDIPTLCFLKDINEVGACRMCVVEVKGARS
15 LQAACVYPVSEGLEVYTQTPAVREARKVTLELILSNHEKKCLTCVRSENCCELQRLAK
DLNVKDIRFEGEMSNLPIDDLSPSVVRDPNKCVCVLCRRCVSMCKNVQTVGAIDVTERG
FRTTVSTAFNKLSEVPCVNCGQCINVCVPGALREKDDIDKVWEALANPELVVVQT
APAVRV ALGEEFGMPIGSRVTGKMVAALSRLGFKKVFDTDTAADLTIMEEGTELINR
IKNGGKLPLITSCSPGWIKFCEHNYPEFLDNLSSCKSPHEMFGAVLKSYYAQKNGIDP
20 SKVFGVSIMPCTAKKFEAQRPELSSTGYPDVDVLTRELARMIKETGIDFNSLPDKQ
FDDPMGEASGAGVIFGATGGVMEAAIRTVGELLSGKPADKIEYTEVRGLDGIKEASIE
LDGFTLKA AVAHGLGNARKLLDKIKAGEADYHFIEIMACPGGCINGGGQPIQPSSVR
NWKDIRCERAKAIYEEDESLPIRKSHENPKIKMLYEEFFGEPGSHKAHELLHTHYEKR
ENYPV

25

SEQ ID NO 95: B. thetaoimicron gi:29345534 NP_809037.1

MEEKQITLQIDGHFITVPEGSTILEAACKIGINIPTLCHIDLKGTCTIKNNPASCRICVVEV
30 AGRRLAPACATRCTEGMVVKTSSTLRVMNARKVVAELILSDHPNDCLTCPKCGNCE
LQTLALRFNIREMPFNGGELSPRKREVTSSIVRNMDKCIFCRRCESVCNDVQTVGALG
AIRRGFNTTIAPAFDRMMKDSECTYCGQCVAVCPVGALTERDYTNRLDDDLADPDKI
VIVQTAPAVRAALGEEFGLPPGTLVTGKMVYALRELGFYVFDTDFAADLTIMEEGS
EILNRLTRYLDGDKSVRLPILTSCCPAWVNFFEHHPDMLDIPSTARSPQQMFGSIAKS

YWAEMGIPREKL VVVSIMPCCLAKKYECDRDEFK VNGVPD VDISISTRELARLIKRA
NIGFTLVLDSPFDNPMGESTGAGVIFGTTGGVMEAAALRSVYEIYTGQPLKNVNFQV
RGLSGVRRATIDLNGFELKVGLAHGLGNARHLLDIRNGHNEYHVIEIMACPGGCIGG
GGQPLHHGNSDVLYARANALYREDANKPLRKSHDNPIYQKLYEEYL GKPLGEKSEM
5 LLHTHYFNKSID

SEQ ID NO 96: gi:2126449 D57150 D. fructosovorans

10 MSMLTTITIDGKTTSVPEGSTILDAAKTLDIDIPTLCYLNLEALSINNKAASCRVCVVEV
EGRRNLAPSCATPVTDNMVVKTNLSRLVNLNARRTVLELLLS DHPKDCLVCAKSCECE
LQTLAERFGIRES PYDGGEMSHYRKDISASIIRDMDKCIMCRR CETMCNTVQTCGVLS
GVNRGFTAVVAPAFEMNLADTVCTNCGQCVAVCPTGALVEHEYIWEVVEALANPD
15 KVVIVQTAPAVRAALGEDLG VAPGTSVTGKMAAALRRLGFDHVFDTDFAA DLTIME
EGSEFLDRLGKHLAGDTNVKLPILTSCCPGWVKFFEHQFPDMLDVPSTAKSPQQMFG
AIAKTYADLLGIPREKL VVVSVMPCCLAKKYECARPEFSVNGNPDVDIVITTRELAKL
VKRMNIDFAGLPDEDFDAPLGASTGAAPIFGVTGGVIEAALRTAYELATGETLKKVD
FEDVRGMDGVKKAKVKVGDNELVIGVAHGLGNARELLKPCGAGETFHAIEVMACP
20 GGCIGGGGGQPYHHGDVELLKKRTQVLYAEDAGKPLRKSHENPYIIELYEKFLGKPLS
ERSHQLLHTHYFKRQRL

SEQ ID NO 97: D. vulgaris gi:97381 S13526

25 MNAFINGKEVRCEPGR TILEAARENGHFIPTLCELADIGHAPGT CRVCLVEIWRDKEA
GPQIVTSC TTPVEEGMRIFTRTPEVRRMQRLQVELLLADHDHDC AACARHGDCELQ
DVAQFVGLTGTRHHFPDYARSRTDVSSPSVVRDMGKCIRCLRCVAVCRNVQGV D
ALVVTGNGIGTEIGLRHNRSQSASDCVGCGQCTLVCPVGALAGRDDVERVIDYLYDP
EIVTVFQFAPAVRVGLGEEFGLPPGSSVEGQVPTALRLLGADVLD TNFAADLVIME
30 EGTELLQRLRGGA KLPLFTSCCPGWVNF AEKHLDPILPHVSTTRSPQQCLGALAKTY
LARTMNVAPERMRVVS LMPCTAKKEEAARPEFRRDGVRD VDAVLTTREFARLLRRE
GIDLAGLEPSPCDDPLMGRATGA AVIFGTTGGVMEAAALRTVYHVLNGKELAPVELH
ALRGYENVREAVVPLGEGNGSVKVAVVHGLKAARQMVEAVLAGKADHVFVEVM

ACPGGCMDDGGGQPRSKRAYNPNAQARRAALFSLDAENALRQSHNNPLIGKVYESFL
GEPCSNLSHRLHLHTRYGDRKSEVAYTMRDIWHEMTLGRRVRGDS

5

SEQ ID NO 98: *T. vaginalis* gi:11127701 AAG31037.1

ASTGINSTANILRNITVTVNGKPLEAKKGETVLELCDRNNIRIPRLCFHPNLPPKASCR
10 VCLVECDGKWLSPACVTTVWDGLKIDTKSKNVRDSVENNLKELLDCHDETCSACIA
NHRCQFRDMNVAYSVKAETKEICSEEGIDESTNAIRLDTSKCVLCGR CIRACEEVAGT
SAIFGNRAKKMRIQPTFGVTLQETSCIKCGQCTLYCPVGAITEKSQVKEALDILANKG
KKITVVQVAPAVRVALSEAFGYKEGTVTTGKMVSALKALGFDLVYDTNYGADLTIC
EEAGELVNRLRDPNAKFPMTTCCPAWVNYVEQSAPDFIPNLSSCRSPQGMLSALIK
15 NYLPKLLDVKQEDVLNFSIMPCTAKKDEVERPELRTKSGLKETDMVLTVRELVEMIK
LSNIDFNNLPDTQFDNIFGFGSGAGQIFAATGGVMEASRTAFEVYTGKKLTNVNIYP
VRGMDGLRIAELDLGDKLKVAVCHGIANTAKLLDRLREKDPPELMDIKFIEIMACPG
GCVCGGGTPQPKNRVSLDNRLAAIYNIDAKMECRKSHENPLIKGVYKEFLGKPNSHL
AHELLHTHFKHHPKW

20

SEQ ID NO 99: gi:7522122.T18557

MISRLIAKKAPLFLRTFATSEMIŠLKIDGKIISVPKGIMLADAIKKAGANVPTMCYHPD
LPTSGGICRVCLVESAKSPGYPIISCRTPVEEGMEIVTQGSKMKEYRQANLALMLSRH
PNACLSCTSNNTCKTQELSANMNIGQCGFANATPPKNDDSYDMTTAJERDNDKCINC
25 DICVHTCSLQGLNALGFYNEEGHAVKSMGTLDVSECIQCGQCINRCPTGAITEKSEIR
PVLDAINIQQRLVFQMAPSIRVAVAEFGIKPGEKILKNEIATALRKLGSNVFVLDTNF
SADLTIEEGHELIERLYRNVGTGKKLLGGDHMPIDLPMLTSCCPGWIMFIEKNYPDLL
NNLSTCKSPQGM LGALIKGYWAKNIKKMDPKDIVSVSIMPCTAKKAEKERPQLRGD
EGYKDVDYILTTRELAKMLKQSNIDLAKMEPTPFDKVMSEGTGA AVIFGVTGGVME
30 AALRTANEVITGREVPFKNLNIEAVRGM EGIREAGIKLENVLDKYKAFEGVTVKVAI
AHGPNNAKRVMDIHKQAKESGKPAPWHFVEVMACPGGCIGGGGQPKPTNLEIRQAR
TQLTFKEDMDLPLRKSHDNPEIKAIYENYLKEPLGHNSHHYLHTTYSSQKVRDMNLY
NANEAAGLDEILAKYPKEKEYLMPIIIIEHDKKGYISDPSIVKISEHLGMYP AQIESILS
SYHYFPREHTIAILMSICVHCHNCMMKGQGRLKTIQETYDIHETHGGVAKDGSFTL

HTLNWLGYCVNDAPAMMIKRKGTNYVETFTGLLGDNIDQRLKSLKNLKKELPKWP
KNNIREMKSQRNGNSYSCMNTQAPIAEATKKAVSMGPEKVIEEVFKSNLVGRGGAG
FRTGKKWESAYKTPASDKYVVCNADEGLPSTYKDWCLLNNEAKRKEVFTGMGICA
KTIGAKRCFMYLRYEYRNLVPALEQSIKDVSQSTCPELADLKYEIRLGGGPYVAGEEN
5 AQFESIEGRAPLPRKDRPGNIFPTMEGLFHKPTVINNVETFFAIPHIIQQGSQSFGEGKM
PKLLSVTGDVDEPILIEITNLNNYSLNHLLQEISAKDIVAAEIGGCTEPIIFGSKFDTLFGF
GRGTLNAVGSVVLFNSSCDLGKIYENKLFMAEESCKQCVPCRDGSYIFHRAFKELR
DTGKSSYNMRALAVASESAARSSICAHGKALESFLKSACDFMNKTKPIYQPHSTYHQ

10

SEQ ID NO 100: *T vaginalis* gi:1171117 AAC47159.1

MLASSATAMKGFANSLRMKDYSSTGINFDMTKCINCQSCVRACTNIAQONVLKSLT
VNGKSVVQTVTGKPLAETNCISCGQCTLGCPKFTIFEADAINPVKEVLTKKNGRIAVC
15 QIAPAIRINMAEALGVPAGTISLGKVVTALKRLGFDYVFDTNFAADMTIVEEATELVQ
RLSDKNAVLPMTSCCPAWVNYVEKSDPSLIPYLSSCRSPMSMLSSVIKNVFPKKIGT
TADKIYNVAIMPCTRKKDEIQRSQFTMKDGKQETGAVLTSRELAKMIKEAKINFKEL
PDTPCDNFYSEASGGGAIFCATGGVMEAAVRSAYKFLTKKELAPIDLQDVRGVASG
VKLAEVDIAGTKVKVAVAHGIKNAMTLIKKIKSGEEQFKDVKFVEVMACPGGCVVG
20 GGSPKAKTKKAVQARLNATYSIDKSSKHRTSQDNPQLQLYKESFEGKFGGHVAHH
LLHTHYKNRKVNP

SEQ ID NO 101: *C. acetobutylicum* gi:15893326 NP_346675.1

25 MKTIILNGNEVHTDKDITILELARENNDIPTLCFLKDCGNFGKCGVCMVEVEGKGF
RAACVAKVEDGMVINTESDEVKERIKKRVSMMLDKHEFKCGQCSRRENCEFLKLVI
KTKAKASKPFLPEDKDALVDNRSKAIVIDRSKCVLCGRCVAACKQHTSTCSIQFIKID
GQRAVGTVDVCLDDSTCLLCGQCVIACPVAALKEKSHIEKVQEALNDPKKHVIVA
MAPSVRTAMGELFKMGYGKDVTGKLYTALRMLGFDKVFDINFGADMTIMEEATEL
30 LGRVKNNGPFPMTSCCPAWVRLAQNYHPELLDNLSSAKSPQQIFGTASKTYPPSISG
IAPEDVYTVTIMPCNDKKYEADIPFMETNSLRDIDASLTRELAKMIKDAKIKFADLE
DGEVDPAMGTYSAGAGIFGATGGVMEAAIRSAKDFAEENKELENVDYTEVRGFKGIK
EAEVEIAGNKLNVAVINGASNFFEFMKSGKMNEKQYHFIEVMACPGGCINGGGQPH

VNALDRENVDIRKLASVLYNQDKNVLSKRKSHDNPAIIKMYDSYFGKPGEGLAH
KLLHVKYTKDKNVSKHE

5

SEQ ID NO 102: gi130069 P29166 c.p?

MKTHINGVQFNTDEDTTILKFARDNNIDISALCFLNNCNNDINKCEICTVEVEGTGLV
TACDTLIEDGMIINTNSDAVNEKIKSRISQLLDIHEFKCGPCNRRENCEFLKLVIKYKA
10 RASKPFLPKDKTEYVDERSKSLTVDRTKCLLCGRCVNACGKNTETYAMKFLNKNGK
TIIGAEDKCFDDTNCLLCGQCIIACPVAALSEKSHMDRVKNALNAPEKHVIVAMAP
SVRASIGELFNMGFGVDVTGKIYTALRQLGFDKIFDINFGADMTIMEEATELVQRIEN
NGPFFMFTSCCPGWVRQAENYYPELLNNLSSAKSPQQIFGTASKTYYPSSISGLDPKNV
FTVTVMPCTSKKFEADRPQMEKDGLRDIDAVITTRELAKMIKDAKIPFAKLEDSEAD
15 PAMGEYSGAGAIFGATGGVMEAAALRSAKDF AENAELEDIEYKQVRGLNGIKEAEVEI
NNNKYNVA VINGASNLFKFMKSGMINEKQYHFIEVMACHGGCVNGGGQPHVNPKE
LEKVDIKKVRASVLYNQDEHLSKRKSHENTALVKMYQNYFGKPGEGRAHEILHFYK
KK

20 SEQ ID NO 103: gi:130070 P07598 ?

MSRTVMERIEYEMHTPDPKADPKLHFVQIDEAKCIGCDTCSQYCPTAAIFGEMGEP
HSIPHIEACINCGQCLTHCPENAIYEAQSWVPEVEKKLKDGVKCIAMPAPAVRYAL
GDAFGMPVGSVTTGKMLAALQKLGFHCWDTEFTADVTTWEEGSEFVERLTKKSD
25 MPLPQFTSCCPGWQKYAETYYPELLPHFSTCKSPIGMNGALAKTYGAERMKYDPKQ
VYTVSIMPCIAKKYEGLRPELKSSGMRDIDATLTRELAYMIKKAGIDFAKLDPGKR
DSL MGESTGGATIFGVTGGVMEAAALRFAYEAVTGKKPDSWDFKAVRGLDGIKEATV
NVGGTDVKVAVVHGA KRFKQVCDDVKAGKSPYHFIEYMACPGGCVCVGGGQPVMP
GVLEAMDRTTTRLYAGLKKRLAMASANKA

30

SEQ ID NO 104: ? gi:1345094 AAC47160.1

MLASSSRAAANIRWVDTSNIAIAFDMH KCINCQACVRACKNVAGQSVLKS VKINEG
KKKGVVQTVTGKLLAETNCIGCGQCTLVCPTQAIHEKDALKQMNNIFKNKGDRILV

CQIAPAIRINMRRPWCSSRNSFHRQSRYSYPQRLGFDYVFDTNFGADLTIVEEATELLQ
RLNDPKAVLPMFTSCCPAWVNYVEKSYPQWMPHLSTCRSPIGMLSAVIKNVFPKHIG
VDPKRIFSVGIMPCTAKKDEAAREQLMTKSGLHETDLDITSRELAKMIKAAKINFKEL
PDTELDSPYAMATGGGAIFCATGGVMEAAVRSAYKFATGKELAPIEFVQVRGAEKGI
5 KVGTVDINGREIKVAVAQGVKNAMSLIKKIEEGQDDVKGVVFCEVMACPGGCVGG
GGSPRAKTKAAMNKRLDATYRIDRASKYRTPQDNTQLQDLYNATWVVS LVMD

SEQ ID NO 105: gi:15644177 NP_229226.1 T. maritima

10 MKIYVDGREVIINDNERNLLEALKNVGIEIPNLCYLSEASIYGACRMCLVEINGQITTS
CTLKPYEGMKVKTNTPETIYEMRRNILELILATHNRDCTTCDRNGSCKLQKYAEDFGI
RKIRFEALKKEHVRDESAPVVRDTSKCILCGDCVRVCEEIQGVGVIEFAKRGFESVVT
TAFDTPLIETECVLCGQCVAYCPTGALSIRNDIDKLIEALES DKIVIGMIAPAVRAAIQE
15 EFGIDEDVAMAEKLV SFLKTIGFDKVFDVSFGADLVAYEEAHEFYERLKKGERLPQF
TSCCPAWVKHAEHTYPQYLQNLSSVKSPQQALGTVIKKIYARKLGVPEEKIFLV SFM
PCTAKKFEAEREEHEGIVDIVLTTRELAQLIKMSRIDINRVEPQPFDRPYGVSSQAGLG
FGKAGGVFSCVLSVLNEEIGIEKVDVKSPEDGIRVAEVT LKDGTSFKGAVIYGLGKV
KKFLEERKDVEIIEVMACNYGCVGGGGQYPNDSRIREHRAK VLRDTMGIKSL LTPV
20 ENLFLMKLYEEDLKDEHTRHEILHTTYRPRRRYPEKDVEILPVPNGEKRTVKVCLGT
SCYTKGSYEILKKLV DYVKENDMEGKIEVLGTFCVENCGASPNVIVDDKIIGGATFEK
VLEELSKNG

25 SEQ ID NO 106: T vaginalis gi:19547859 AAK26191.1

CDGKWLAPACVTTVWDGLKIDTKSKMVKESVENNLKELLDCHDETCS SCVANHRC
QFRDMNVAYSIIKAETKEECSEEGIDESTNSIRLDTSKCVLCGRCIRACEEVAGQSAIIF
GNRAKHMRIQPTFGQTLQDTSCIKCGQCTLYCPVGAITEKSQVKQALDILSNKGKKIS
30 VIQVAPAVRV ALSEAFGYKEGSVTTGKMVSALKALGFDYVYDTNYSADLTIVEEAG
ELVQRLKNPNAVFPMFTSCCPAWVNYVEQSAPDFIPNLSSCRSPQGMLSSLVKNYLP
KVLNIPVEDVLNFSIMPCTAKKDEIERPELRTKDGHKETDMVLT VREL VEMIKLSGID
FNNLPDTPFDSIFGFGSGAGQIFAAT

SEQ ID NO 107: *R. norvegicus* gi:34870568 XP_340771.1

MASPFSGALQLTDLDDFIGPSQSCIKPVTVAKKPGSGIAKIHIEDDGSYFQVNPDGRSQ
5 KLEKAKVSLNDCLACSGCVTSAETILITQQSHEELRKVLDANKVAAPGQQRLVVVSV
SPQSRASLAARFQLDSTD TARKLTSFFKKIGVHFVFD TAFARNFSLLESQKEFVQRFR
EQANSREALPMLASACPGWICYAEKTHGNFILPYISTARSPQQVMGSLIKDFFAQQQL
LTPDKIYHVTVMPCYDKKLEASRPDFFNQEYQTRDVDCVLT TGEVFRLLLEE EGVSL
ELEPVPLDGLTRSVSAEPTSHRGGGSGGYLEHVFRHAAQELFGIHVADVTYQPMRN
10 KDFQEVTLEREGQVLLRFAVAYGFRNIQNLVQKLKRGRCPYHYVEVMACPSGCLNG
GGQLKAPDTEGRELLQQVERLYSMVRTEAPEDAPGVQELYQHWLQGEDSERASHLL
HTQYHAVEKINSGLSIRW

15 SEQ ID NO 108: *S. cerevisiae* gi:14336719 AAK61251.1

MASPFSGALQLTDLDDFIGPSQVGS LQALLALAFLHTGNFSAAGCWEPDPWECIKPV
KVEKRAGSGVAKIRIEDDGSYFQINQEKLGEELEPTFGIFLPYSPDGGTRRLEKAKVS
LNDCLACSGCITSAETVLITQQSHEELKKVLDANKMAAPSQQRLVVVSVSPQSRASL
20 AARFQLNPTD TARKLTSFFKKIGVHFVFD TAFSRHFSLLESQREFVRRFRGQADCRQA
LPLLASACPGWICYAEKTHGSFILPHISTARSPQQVMGSLVKDFFAQQQHLTPDKIYH
VTVMPCYDKKLEASRPDFFNQEHQTRDVDCVLT TGEVFRLLLEE EGVSLPDLEPAPLD
SLCSGASAEPTSHRGGGSGGYLEHVFRHAAARELFGIHVAEVTYKPLRNKDFQEVTL
EKEGQVLLHFAMAYGFRNIQNLVQRLKRGRCPYHYVEVMACPSGCLNGGGQLQAP
25 DRPSRELLQHVERLYGMVRAEAPEDAPGVQELYTHWLQGT DSECAGRLLHTQYHA
VEKASTGLGIRW

SEQ ID NO 109: *C. perfringens* gi:18311328 NP_563262.1

30 MNKIIINDKTIEFDGDKTILDLARENGFDIPVLCELKNCGNKGQCGVCLVEQEGNDRL
LRSCAIKAKDGMVIKTDSEKVLEARKERVAELLDEHEFKCGPCKRRENCEFLKLVK
TKARAHKPFVVADKSEYVDDRKSIVLDRSKCVKCGRCVAACRTRTATNSIKFHRID
GVRLVGPEELKCFDDTNCLLCGQCIAACPVDALSEKSHIERVQDALNDPEKHVIVAM
APAVRTSMGELFKMGYGQDVTGKLYTALRELGFDKVFDINFGADMTIMEEATELIE

RIKNNGPFPMILTSCCPSWVREVENYFPELVENLSSAKSPQQIFGAASKTYYPQVADID
PKKVFTVTVMPTCKKFEADRPENEGIRNIDAVITTRELARMIKAAKIDFAKLEDG
EVDPAAMGEYTGAGVIFGATGGVMEALRTAKDFMENDNLDNVDYEAVRGLAGIKE
AEVEIAGNEYKLAVVSGAANVFELVKSGKINDYHFIEVMACPGGCVNNGGGQPHISAE
5 DSDKIDIREVRASVLYNQDKNLEKRKSHQNS ALLKMYENYMGKPGHGRAHELLHM
KYKK

SEQ ID NO 110: gi:4239873 BAA74726.1 C. perfringens

10 MNKIIINDKTIEFDGDKTILDLARENGFDIPVLCCKNCGNKGQCGVCLVEQEGNDRL
LRSCAIKAKDGMVIKTDSEKVLEARKERVAELLDEHEFKCGPCKRRENCEFLKLVK
TKARAHKPFVADKSEYVDDRSKSIVLDRSKCVKCGRCVAACTRTATNSIKFHRID
GVRLVGPEELKCFDDTNCLLCGQCIACPVDALEKSHIERVQEALNDPEKHVIVAM
15 APAVRTSMGELFKMGYGGQDVTGKLYTALRELGFQKVDINFGADMTIMEEATELIE
RIKNNGPFPMILTSCCPSWVREVENYFPELVENLSSAKSPQQIFGAASKTYYPQVADID
PKKVFTVTVMPTCKKFEADRPENEGIRNIDAVITTRELARMIKAAKIDFAKLEDG
EVDPAAMGEYTGAGVIFGATGGVMEALRTAKDFMENDNLDNVDYEAVRGLAGIKE
AEVEIAGNEYKLAVVSGAANVFELVKSGKINDYHFIEVMACPGGCVNNGGGQPHISAE
20 DSDKMDIREVRASVLYNQDKNLEKRKSHQNSALLKMYESYMGKPGHGRAHELLH
MKYKK

25 SEQ ID NO 111: C. tetani gi:28212003 NP_782947.1

MIVFENQLKKLKYLVLKEVAKMTLEDRLGEEDIERISFDIHKGDKAERYCCVYKERAI
VYERAKLATGCLPNGQVAEEFVHVEDDDQIYVIDAACDKCPINKYVVTEACRGCLQ
HKCMEVCPAGSINRAAGKAYINHETCKEGLCESACPYNAIAEVMRPCRRACPTGA
30 LQMNLEDNKATINKEDCINCGSCMSVCPFGAISDKSYIVDITKALKNNKKVYAMVAP
AITGQFGKDVSVGKMKNAFKAMGFEDMLEVACGADAVAAHESEEFIERLESGKKY
MTTSCCPGFLGYIEKKFPDQLENVSNTVSPMVAIGRMIKKEYEDSVVVVFGPCTAKK
AEIKRKGIKDAVDYVMTFEEIAALMGAFEIDPAECEEDINDGSNYGRGFAQGGGVV

SAIQNCIKDKEGIKFNPLRVSGPDQIKRAMIMAKVGKLSNFIEGMMCEGGCIGGPAT
MVSAVKAKAPLMKFSKSSSTIKDVKDNEVLDDKYKDINMER

SEQ ID NO 112: *C. tetani* gi:28209953 NP_780897.1

5

MHNDYREIFKRLSKSYDDTFEKEVENILSSHSMREKLAKIISILCGVNIEHSENYIS
NLKNAIKNYTASAEKVVTKLPCSTQCAKDGDIIICEKSCPVNAIFRDPNDNNIYINDEL
CLDCGLCVRNCPSPGSILDKKEFIPLAELLKSESIVIAAVAPAIMGQFGENTTINQLRTAF
KKLGFTDMVEVAFFADMLTLKEAVEYDHFVKDEQDFMITSCCCPMWVGMLKKVY
10 NDLVKYVSPSVSPMIAAGRVLKLLNPNCVVVFVGPCIAKKAEAREKDLLGDIDFVLT
FTELRDIFDVFDIQPENLEEDFSSEYASKGGRLYARTGGVSIIVSEAIEKLFPNKKYKFL
KTIQADGVKGCKSLDDKIKQEDISANFVEGMGCVGGCVGGPKVIIDPSEGRNAVNNF
AENSSIKVSVDSNCMNDILSKININSVEDFKDKDKISIFEREFK

15 SEQ ID NO 113: *Pyrococcus furiosus* delta (small) subunit GI:562776 X75255

MGKVRIGFYALTSCYGCQLQLAMMDELLQLIPNAEIVCWFMDRDSIEDEKVDIAFIE
GSVSTEEVELVKKIRENAKIVVAVGACAVQGGVQSWSEKPLEELWKKVYGDAKV
KFQPKKAEPVSKYIKVDYNTYGCPPEKKDFLYALGTFLIGSWPEDIDYPVCLECRNG
20 HPCILLEKGEPCLGPVTRAGCNARCPGFGVACIGCRGAIGYDVAWFDSLAKVFKEKG
MTKEEIIERMKMFNGHDERVEKMVEKIFSGGEQ

SEQ ID NO 114: *Escherichia coli* small subunit GI:2668498 M63654

25

MSPVLTQHVSQPITLDEQTQMKRHLQDIRRSAYVYRVDCGGCNACEIEFAAITPV
FDAERFGIKVVSSPRHADILLFTGAVTRAMRMPALRAYESAPDHKICVSYGACGVGG
GIFHDLYSVWGGSDTIVPIDVWIPGCPPTPAATHGFAVALGLLQKKIHAVDYRDPTG
VTMQPLWPQIPPSQRIAIEREARRLAGYRQGREICDRLLRHLSDDPTGNRVNTWLRD
30 ADDPRLNSIVQQFLFRVLRGLHD

SEQ ID NO 115: *Methanothermobacter thermautotrophicus* GI:149717 small subunit
J02914

MAEENAKPRIGYIHLSGCTGDAMSLTENYDILAELLTNMVDIVYGQTLVDLWEMPE
MDLALVEGSVCLQDEHSLHELKELREKAKLVCAFGSCAQTGCFTRYSRGGQQAQPS
HESFVPIADLIDVDLAIPGCPPSPEIIAKAVVALLNNDMEYLQPMLDLAGYTEACGCD
5 LQTKVVNQGLCTGCGTCAMACQTRALDMTNGRPELNSDRCIKCGICYVQCPRSWW
PEEQIKKELGL

10 SEQ ID NO 116: Y13764 *Methanosarcina barkeri* GI:2463276 small subunit

MANKIKLGHVHLSGCTGCLVSVADNYQGFLKILDDYADLVYCLTLADVRHIPEMDV
ALVEGSVCIQDRESVEDIKETRKKSRIVVALGSCASYGNITRFCRGGQHNPQHESYL
PIGDLIDVDVYIPGCPPSPELIRNVAIMAYLLLEGNEEQKDLAGRYLKPLMDLAKRGT
15 TGCFCDLMDDVINQGLCIGCGICAASCPVRAITHEFGKPQGDNLNLCIKCGSCYGACPR
SFFNPDVISEFESINEIIAGALKEGEKDD

20 SEQ ID NO 117: U65510 *Rhodospirillum rubrum* GI:1515468 small subunit

MNFLSRMSKKSPWLYRINAGSCNGCDVELATTACIPRYDVERLGCQYCGSPKHADI
VLVTGPLTARVKDKVLRVYEEIPDPKVTVAIGVCPISGGVFRESYSIVGPIDRYLPVDV
NVPGCPPRPQAIIIEGIAKAIEIWAGRI
25

SEQ ID NO 118: X75255. *Pyrococcus furiosus* alpha (large) subunit GI:562777

MKNLYLPITIDHIARVEGKGGVEIIIGDDGVKEVKLNIEGPRFFEAITIGKKLEEALAIY
30 PRICSFCSAAHKLTALEAAEKAVGFVPREEIQALREVLYIGDMIESHALHLYLLVLPD
YRGYSSPLKMVNEYKREIEIALKLKNLGTWMMDILGSRAIHQENAVLGGFGKLPEKS
VLEKMKAEALREALPLAEYTFELFAKLEQYSEVEGPITHLAVKPRGDAYGIYGDYIKA
SDGEEFPSEKYRDIKEFVVEHSFAKSHYKGRPFMVGAISRVINNADLLYGKAKEL
YEANKDLLKGTPFANNLAQALEIVYFIERAIDLLDEALAKWPIKPRDEVEIKDGFV

STTEAPRGILVYALKVENGRVSYADIITPTAFNLAMMEEHVRMMAEKHYNDPPERL
KILAEMVVRAYPDPCISCSVHVRL

5

SEQ ID NO 119: M63654. *Escherichia coli* large subunit GI:1616956

MNVNSSSNRGEAILAALKTQFPGAVLDEERQTPEQVTITVKINLLPDVVQYLYYQHD
GWLPLVLFGNDERTLNGHYAVYYALSMEGAECWIVVKALVDADSREFPSVTPRVP
10 AAVWGEREIRDMYGLIPVGLPDQRRLVLPDDWPEDMHPLRKDAMDYRLRPEPTTDS
ETYPFINEGNSDARVIPVGPLHITSDEPGHFRLFVDGEQIVDADYRLFYV
HRGMEKLAETRMGYNEVTFLSDRVCGICGFAHSVAYTNSVENALGIEVPQRAHTIRS
ILLEVERLHSHLLNLGLSCHFVGFDTGFMQFFRVREKSMTMAELLIGSRKTYGLNLIG
GVRRDILKEQRLQTLKLVREMRADVSELVEMLLATPNMEQRTQGIGILDRQIARDLR
15 FDHPYADYGNIPKTLFTFTGGDVFSRVMVRVKETFDLAMLALDNPDPDTPLLTEG
FSYKPHAFALGFVEAPRGEDVHWSMLGDNQKLFRWRCRAATYANWPVLRVYMLRG
NTVSDAPLIIGSLDPCYSCTDRVTLVDVRKRQSKTVPYKEIERYGIDRNRSPK

20 SEQ ID NO 120: J02914. *Methanothermobacter thermautotrophicus* large subunit
GI:551889

MSERIVISPTSRQEGHAELVMEVDDEGIVTKGRYFSITPVRGLEKIVTGKAPETAPVIV
QRICGVCPIPHTLASVEAIDDSLDIEVPKAGRLLRELTAAHHVNSHAIHHFLIAPDFV
25 PENLMADAINSVSEIRKNAQYVVDMMVAGEGIHPSDVRIIGGMADNITELARKRLYARL
KQLKPKVDEHVELMIGLIEDKGLPKGLGVHNQPTLASHQIYGDRTKFDLDRFTEVMP
ESWYDDPEIAKRACSTIPLYDGRNVEVGPRARMVEFQGFKERGVVAQHVARALEM
KTALARAIEILDELDTAPVRADFDERGTGKLGVGAIIEGPRGLDVHMAQVENGKIQF
YSALVPTTWNIPMTMGPALEGFHHFYGPHVIRAYDPCLSCATHVMVVDDEDRSVIRDE
30 MVRL

SEQ ID NO 121: Y13764 *Methanosarcina barkeri* GI:2463274 large subunit

MTKVVEISPTTRHEGHSKLTCLKVNDEGIVERGDWLSTTPVRGIEKLAIGKTMDQVPKI
ASRVCGICPIAHTLAGIEAMEASIGCEIPKDAKLLRVILHAANRLHSHALHNILILPDFY
IPDTETKINPFSKEQPLRSVAVRIFRIREIAQTIGAVAGGEAIHPSNPRVGGMYRNVSSR
AKQKIADLAKEGLVLAHEQMEFMIEVIRNMQDREFVEVAGKQIPLPKTLGYHNQGV
5 MATAPMYGSSSLDEKPMWDFTRWRETRPVDWYMSEETIDLED
SSYPIGGTTKVGTKVNPRMEACNTVPTYDGGQPVVGPRARLATFKHFTEKGTFAQHI
ARQMEYTDCCYYTILNCLNLDTSKGVLADTIPLGNGSMGWAANEAPRGTDVHLAR
VKDGKVLRYEMLVPTTWNFPTCSRALTGAPWQIAEMVIRAYDPCVSCATHMIVVNE
EDRIVAQKLMQW

10

SEQ ID NO 122: U65510 *Rhodospirillum rubrum* large subunit GI:1498746

MSTYTIPVGPLHVALEPPMYFRIEVDGEKVVSVSDITAGHVHRGIEYLATKRNIYQNIV
15 LTERVCSLCSNSHPQTYCMALESITGMVPPRAQYLRVIADETKRVASHMFNVAILA
HIVGFDSLFMHVMEAREIMQDTKEAVFGNRMDIAAMAIGGVKYDLDDKDGRDYFIG
QLDKLEPTLRDEIPLYQTNPSIVDRTRGIGVLSAADCVDYGLMGPVA
RSGSHAYDVRKQAPYAVYDRLDFEMALGEHGDVWSRAMVRWQEALTSIGLIRQCL
RDMPDGPTKAGPVPPIPAGEAVAKTEAPRGELIYYLKTNGTDRPERLKWRVPTYMN
20 WDALNVMMAGARISDIPLIVNSIDPCISCTER

SEQ ID NO 123: *Chlamydomonas reinhardtii* background with higher molecular weight
variant amino acid (naturally occurring in *D. vulgaris*) (TIMEE→TIWEE)

25 MALGLLAELRAGQAVACARRTNAPAHAAVVPCLPSRAGKFFNLSQKVPSSQSARG
STIRVAATATDAVPHWKLAEELDKPKDGGKVLIAQVAPAVRVAIAESFGLAPGA
VSPGKLATGLRALGFDQVFDTLFAADLTIWEEGTELLHRLKEHLEAHPHSDEPLPMF
TSCCPGWVAMMEKSYPELIPFVSSCKSPQMMMAMVKTYLSEKQGIPAKDIVMVS
MPCVRKQGEADREWFCVSEPGVRDVDHVITTAELGNIFKERGINLPELPDSDWDQPL
30 GLGSGAGVLF GTTGGVMEALRTAYEIVTKEPLPRLNLSEVRGLDGIKEASVTLVPA
PGSKFAELVAERLAHKVEEAAAAEAAA AVEGAVKPPIAYDGGQGFSTDDGKGGLKL
RVAVANGLGNAKKLIGKMVSGEAKYDFVEIMACPAGCVGGGGQPRSTDQITQKR
QAALYDLDERNTLRRSHENEAVNQLYKEFLGEPLSHRAHELLHHTHYVPGGAEADA

SEQ ID NO 124: GAGVIFGATGGVMEALRT
SEQ ID NO 125: GGGAIFCATGGVMEAAVRS
SEQ ID NO 126: GGATIFGVTGGVMEALRF
SEQ ID NO 127: GAGAIFGATGGVMEALRS
5 SEQ ID NO 128: GAGAIFGATGGVMEAAIRS
SEQ ID NO 129: GAAVIFGVTGGVMEALRT
SEQ ID NO 130: GAGQIFAATGGVMEASRT
SEQ ID NO 131: GAAVIFGTTGGVMEALRT
SEQ ID NO 132: GAAPIFGVTGGVIEAALRT
10 SEQ ID NO 133: GAGVIFGTTGGVMEALRS
SEQ ID NO 134: GAGVIFGATGGVMEAAIRT
SEQ ID NO 135: SAGNLFVTGGVMEAAIRT
SEQ ID NO 136: GAGAIFGATGGVMEALRT
SEQ ID NO 137: GAGVLFGTTGGVMEALRT
15 SEQ ID NO 138: GAAALFGVTGGVMEALRT
SEQ ID NO 139: GAGVLFGTTGGVMEAAVRT
SEQ ID NO 140: GAGTIFGTTGGVMEALRT
SEQ ID NO 141: GGGVLFGTTGGVMEALRT
SEQ ID NO 142: TIMEE
20 SEQ ID NO 143: TIVEE
SEQ ID NO 144: TIWEE
SEQ ID NO 145: TICEE
SEQ ID NO 146: VIMEE
SEQ ID NO 147: TARLE

25

SEQ ID NO 148: promoter of the lhcb1 gene: Genbank accession number Y16833:
gcagttgggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggccaacaagtcgcctcgggtcgctgtcggc
gccaaccgcagctgcatccaccagattcacttgtagatcgacctagggtgcgggaccggaggcggctcgctgtgcaagcgcggt
gacctgtacggcgcatggatcgccatctcgattcgcgcggcagaatcgggcccgcgcacatttaagccgcgggcgagactcat
30 ttcgtta

SEQ ID NO 149: Selectable marker: (Lumbreras, Plant J (1998) 14(4): 441-447)
gccagaaggagcgcagccaaaccaggatgatgttgatggggtatttgagcacttgcaaccctatccgggaagccccctggcccaca
aaggctaggcgcgaatgcaagcagtcgcatgcagcccctggagcgggtgccctcctgataaacggccaggggcctatgttctta

ctttttacaagagaagtcaactcaacatctaaaatggccaggtgagtcgacgagcaagcccggcgatcaggcagcgtgctgcaga
tttgacttgcaacgcccgcattgtgtcgcgaaggcttttggtcctctgtcgtgtcgaagcagcatctaaccctgcgtgccgtttcca
tttgaggatggccaagctgaccagcgccgttccggtgtcaccgcgcgcgacgtgccggagcggtcgagttctggaccgaccgg
ctcgggttctccgggacttcgtggaggacgacttcgccggtgtggtccgggacgacgtgacctgttcacagcgcggtccaggac
5 caggtgagtcgacgagcaagcccggcgatcaggcagcgtgctgcagattgacttgcaacgcccgcattgtgtcgcgaaggctt
ttggtcctctgtcgtgtcgaagcagcatctaaccctgcgtgccgtttccattgcaggaccagggtgtgccggacaacacctggc
ctgggtgtgggtgcgcggcctggacgagctgtacgccgagtggtcggaggctgtccacgaacttcgggacgacctccgggccc
gccatgaccgagatcgccgagcagccgtggggcgggagttcgccctgcgcgacccggccggcaactgcgtgacttcgtggcc
gaggagcaggactaacgacgtcgaaccactctagaggatcgaiccccgctccgtgtaaatggaggcgctcgttgatctgagccttg
10 cccctgacgaacggcggtggatggaagatactgctcgaagtgtgaagcggtagcttagctccccgttctgctgacagctctttt
caacacgtaaaaagcggagagtttgcaattttgttggtgtaacgatcctccgttgattttggcctctttccatggcggggctgggcg
tatttgaagcttaattaactcaggggggggcccgtacc

15 SEQ ID NO 150: 3' untranslated region of RBCS2 gene

ccgacgtcgaccactctagaggatcgatccccgctccgtgtaaatggaggcgctcgttgatctgagccttgccccctgacgaacggc
ggtggatggaagatactgctcgaagtgtgaagcggtagcttagctccccgttctgctgacagctcttttcaacacgtaaaaagcg
gaggagtttgcaattttgttggtgtaacgatcctccgttgattttggcctctttccatggcggggctggcggtatttg

20

SEQ ID NO 151: 3' untranslated region of RBCS2 gene - promoter of the lhcb1 gene

ccgacgtcgaccactctagaggatcgatccccgctccgtgtaaatggaggcgctcgttgatctgagccttgccccctgacgaacggc
ggtggatggaagatactgctcgaagtgtgaagcggtagcttagctccccgttctgctgacagctcttttcaacacgtaaaaagcg
gaggagtttgcaattttgttggtgtaacgatcctccgttgattttggcctctttccatggcggggctggcggtatttggcagttgggtca
25 ggggctggcgacgcgctgctgacgcgaagtgaatggcccaacaagtcgcctcgcggtcgtgctggcgccaaaccgcagctgc
atccaccagattcacttgtagatcgacctaggttgcgggaccggaggcggtcgtgtgcaagcgcggtgacctcgtacggcgcca
tggatcgccatctcgattcgcgcggcagaatcgggccccgcgcacatttaagccggggcgagactcatttcgtta

30

SEQ ID NO 152: 5 Unique sequence a: 5' atccgtagttatccttatggccatcttagc 3'

SEQ ID NO 153: Unique sequence b: 5' cgtgcatcgattaacagcttctggacctga 3'

SEQ ID NO 154: Unique sequence c: 5' taaacgtcgtagctcaagtataactaag 3'

SEQ ID NO 155: Unique sequence d: 5' aatctgatacatgctattcagaatcttaca 3'

SEQ ID NO 156: Unique sequence e: 5' tcttccatcgtaaacttagcatcgattagc 3'

5 SEQ ID NO 157: Unique sequence f: 5' atctgtaataatctagtcgaggcattcaag 3'

SEQ ID NO 158: Unique sequence g: 5' aactggctaaatcgtaacaatcgtgtga 3'

SEQ ID NO 159: Unique sequence h: 5' gatttaacataactgtcgattaccgtgcga 3'

10

SEQ ID NO 160: Unique sequence i: 5' tatgcttgacaatcgtaacctggtgacaa 3'

SEQ ID NO 161: Unique sequence j: 5' taacaagaatctggctaataatcgatgca 3'

15 SEQ ID NO 162: Unique sequence k: 5' gtagtcggaatagttactaacgaggattcg 3'

SEQ ID NO 163: Unique sequence l: 5' aaatgtctactcgactagtaaatcgtact 3'

SEQ ID NO 164: Nonshuffled segment I:

20 gcagtgggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggcccaacaagtcgcctcgcggctcgtgtcggc
gccaaaccgcagctgcatccaccagattcacttgttagatcgacctagggtgcgggaccggaggcggctcgtgtgcaagcgcgg
gacctcgtagggcgcatggatcgccatctcgattcgcggcgagaatcgggccccgcgcacatttaagccgaggcgagactcat
ttcgtaatccgtagttatccttatggccatcttagc

25 SEQ ID NO 165: Nonshuffled segment II:

cgatgcatgattaacagcttctggacctgaccgacgtcgacccactctagaggatcgatccccgctccgtgtaaatggaggcgtcgtt
gatctgagccttgcctccctgacgaacggcggtgatggaagatactgctctcaagtgtgaagcggtagcttagctccccgttcgtgc
tgatcagctcttttcaacacgtaaaagcggaggagtttgaatttgggtgtaacgatcctcgttgatttggcctcttctccatggg
cgggctggcggtatttggcagttgggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggcccaacaagtcgcctcgcg
30 gtcgctgctggcgccaaaccgcagctgcatccaccagattcacttgttagatcgacctagggtgcgggaccggaggcggtcgtgt
gcaagcgcgggtgacctgtagggcgcatggatcgccatctcgattcgcggcgagaatcgggccccgcgcacatttaagccgcg
ggcgagactcatttcgttattaaacgtcgtacgtccaagtatgactaag

SEQ ID NO 166: Nonshuffled segment III:

aatctgatacatgctattcagatcttacaaccgacgtcgacccactctagaggatcgatccccgctccgtgtaaattggaggcgctcgttg
atctgagccttgccccctgacgaacggcgggtggatggaagatactgctctcaagtctgaagcggtagcttagctccccgtttcgtgct
gatcagctcttttcaacacgtaaaaaggaggagttttgcaattttgttggtgtaacgatcctccgttgattttggcctcttttccatgggc
gggctgggcgtattttggcagttgggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggcccaacaagtcgctcgcggt
5 cgctgicggcgccaaaccgcagctgcatccaccagattcacttgttagatcgacctagggttgcgggaccggaggcggtcgcgtgtg
caagcgcggtgacctcgtacggcggcatggatgccatctcgattcgcgcggcagaatcgggccccgcgcacatttaagccgcgg
cgatcttccatcgtaaattctagcatcgattagc

SEQ ID NO 167: Nonshuffled segment IV:

10 atctgtaataatctagtcgaggcattcaagccgacgtcgacccactctagaggatcgatccccgctccgtgtaaattggaggcgctcgtt
gatctgagccttgccccctgacgaacggcgggtggatggaagatactgctctcaagtctgaagcggtagcttagctccccgtttcgtgc
tgatcagctcttttcaacacgtaaaaaggaggagttttgcaattttgttggtgtaacgatcctccgttgattttggcctcttttccatggg
cgggctgggcgtatttg

15 SEQ ID NO 168: Nonshuffled segment V:

gccagaaggagcgcagccaaaccaggatgatgttgatggggtatttgagcacttgcaacccctatccggaagccccctggccaca
aaggctaggcgccaatgcaagcagttcgatgcagccccctggagcgggtgccctctgataaacggccagggggacctatgttcttta
ctttttacaagagaagtcactcaacatcttaaaatggccagggtgagtcgacgagcaagcccgcggaatcaggcagcgtgcttgaga
tttgacttgcaacgcccgcattgtgtcgacgaaggcttttggtcctctgtcgtgtctcaagcagcatctaacctcgctgcggtttcca
20 ttgcaggatggccaagctgaccagcgggttccgggtgctcaccgcgcgacgtcgccggagcgggtcgagttctggaccgaccgg
ctcgggttctccgggacttctgtgaggacgacttcgggtgtgttcgggacgacgtgacctgttcatcagcgcggtccaggac
cagggtgagtcgacgagcaagcccgcggaatcaggcagcgtgcttgagatttgacttgcaacgcccgcattgtgtcgacgaaggctt
ttggctcctctgtcgtgtctcaagcagcatctaacctgcgtcgccgtttccatttgaggaccagggtggcggaacacacctggc
ctgggtgtgggtgcgcggcctggacgagctgtacgccgagtggtcggaggtcgtgtccacgaacttcgggacgctccgggccc
25 gccatgaccgagatggcgagcagccgtggggggggagttcgccctgcgcgaccggcgcgcaactgcgtgacttcgtggcc
gaggagcaggactaacgacgtcgacccactctagaggatcgatccccgctccgtgtaaattggaggcgctcgttgatctgagccttg
ccccctgacgaacggcgggtggatggaagatactgctctcaagtgtgaagcggtagcttagctccccgtttcgtgctgatcagcttttt
caacacgtaaaaaggaggagttttgcaattttgttggtgtaacgatcctccgttgattttggcctcttttccatggcggggctgggcg
tattgaagcttaattaactcgagggggggccccgtacc

30

SEQ ID NO 169: Nonshuffled segment VI:

gcagttgggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggcccaacaagtcgacctcgcggtcgtctcggc
gccaaaccgcagctgcatccaccagattcacttgttagatcgacctagggttgcgggaccggaggcggtcgcgtgtgaagcgcggt

gaccctgtacggcggcatggatcgccatctcgattcgcgcgagaaatcgggcccgcgacatttaagccgaggcgagactcat
ttcgtaaaactggcctaaatcgtaacaatcgtgtga

SEQ ID NO 170: Nonshuffled segment VII:

5 gatttaacataactgtcgattaccgtgcgaccgacgtcgaccactctagaggatcgatccccgctccgtgtaaatggaggcgctcgtt
gatctgagccttgccccctgacgaacggcggtggatggaagatactgctctcaagtgtgaagcggtagcttagctccccgtttcgtgc
tgatcagctcttttcaacacgtaaaaaaggaggagtttgaattttgttggtgtaacgatcctccgttgatttggcctcttttccatggg
cgggctgggcgtatttggcagttgggtcaggggctggcgacgcgctgctgacgcgaagtgaatggcccaacaagtcgctcgcg
gtcgtgctggcgccaaacccgcagctgcatccaccagattcactgttagatcgacctaggttgcgggaccggaggcggtcgtgt
10 gcaagcgcggtgacctcgatggcggcatggatcgccatctcgattcgcgcgagaaatcgggcccgcgacatttaagccgcg
ggcgatatgcttgacaatcgaatcctggtgacaa

SEQ ID NO 171: Nonshuffled segment VIII:

taacaagaatctggctaatcaatcgaatgaccgacgtcgaccacictagaggatcgatccccgctccgtgtaaatggaggcgctcgtt
15 gatctgagccttgccccctgacgaacggcggtggatggaagatactgctctcaagtgtgaagcggtagcttagctccccgtttcgtgc
tgatcagctcttttcaacacgtaaaaaaggaggagtttgaattttgttggtgtaacgatcctccgttgatttggcctcttttccatggg
cgggctgggcgtatttgc

SEQ ID NO 172: Chlamydomonas reinhardtii ferredoxin; Genbank accession number

20 L10349
atggccatggctatgcgtccaccttcgccgccggttgccgctaagcccgtgtccggtgctcgccccgccagccgcatgagc
tgcatggcctacaaggtcaccctgaagacccctcgggcgacaagaccattgagtgccccgtgacacctacatcctggacgtgct
gaggaggccggcctggacctgccctactcttgccgcgtggtgcttgcctcagctgcgcggcaaggctgctgccggcaccgtcg
accagtcggaccagtcctcctggacgatgccagatgggcaacggctcgtgctgacctgcgtggcctacccacctcggactgca
25 ccatccagaccaccaggaggaggccctgtactaa

SEQ ID NO 173: Chlamydomonas reinhardtii hydrogenase; Genbank accession number

AF289201

atgtcggcgctcgtgctgaagccctgcgcggccgtgtctattcgcgagctcctgcagggcgcgaggtcgcccccgcgctcc
30 gctcgcagccagcaccgtgcgtgtagcccttgaacacttgaggcgcccgacggcgcttaggcaacgtcgcttgcgcggctgccg
caccgctgcggaggcgctttgagtcagtccagcaggcgctcgccgagcttgccaagcccaaggacgacccacgcgcaagca
cgtcgcgtgcaggtggctccggcggttcgtgctgctattgccgagacctgggctggcgccggggcgccaccaccccaagcagc
tggccgagggccctccgcccctcggttgacgaggtgttgacacgctgtttggcgccgacctgacctcatggaggaggcgagc
agctgctgacccgcctcaccgagcacctggaggccacccgacctccgacgagccgctgcccattgaccagctgctgccccggc

tggatcgctatgctggagaaatctacccggacctgatccctacgtgagcagctgcaagagccccagatgatgctggcgccatg
gtcaagtcttacctagcggaaaagaagggcatcgcgccaaaggacatgggtcatgggtgcatcatgccctgcacgcgcaagcagtc
ggaggtgaccgcgactggttctgtgtggacgccgacccaccctgcgccagctggaccacgtcatcaccaccgtggagctgggca
acatcttcaaggagcgcggcatcaacctggccgagctccccgagggcgagtgggacaatccaatggcggtgggctcggcgccg
5 gcggtgctgttcggcaccaccggcggtgtcatggaggcgcgctgcgcacggcctatgagctgttcacgggcacgccgtgccgcg
cctgagcctgagcgaggtgcgcggcatggacggcatcaaggagaccaatcacatgggtgccgcgccgggtccaagttgag
gagctgctgaagcaccgcgccgccgcgcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc
gcgggcttcaccagcgaggacggcagggggcggcatcacactgcgcgtggccgtggccaacgggctgggcaacgccaagaagct
gatcaccaagatgcaggccggcgaggccaagtacgacttggtagatcatggcctgccccgcccgtgtgtggcgggcgccgcg
10 cagccccgtccaccgacaaggccatcacgcagaagcggcagggcgcgctgtacaacctggacgagaagtccacgtgcgcgcg
agccacgagaaccgtccatccgcgagctgtacgacacgtacctcgagagccgctgggccacaaggcgcacgagctgtgcac
accactacgtggccggcgcgctggaggagaaggacgagaagaagtga

SEQ ID NO 174: *Clostridium pasteurianum* hydrogenase; Genbank accession number

15 M81737

atgaaaacaataattataaatgggttacagtttaatactgatgaagacactactatataaaattgcacgagacaacaattgatatact
gcactgtgtttttaataattgtaataatgacataaataagtgaaatatgtactgtagaggtagagggtactggattagtaacgcctgt
gatacatattgaggatggtatgattataaacacaaattccgatgctgtcaacgaaaaaattaaatctagaatatctcaattattagacata
catgaattcaaatgtgtccttgcaatagaagagaaaactgtgaattctaaaactgttataaaatataaagcaagagctttaaaccattt
20 ttacctaaagataagactgaatatgtatgaaagaagtaaatcattactgtatagtaggacaaaatgcttattatgtgaagatgtgttaa
tgcctgtggaaaaatactgaaacctatgcaatgaaattttaacaaaaatggtaaaactataattggagcagaggatgaaaaatgcttt
gatgataactaattgtctattatgtgtcaatgtataatcgctgtccagtagcagcattatcggaataacacacatggatagagtaaaaa
tgccttaaatgcccctgaaaaacatgtaatagtatggtccatctgtcagagcttctataggtgaacttttaatatgggatttgcgtt
gacgtaacaggaaaaatttatactgcttaagacagcttgatttgataaaatattcgatataaacttcggagcagatatgacaattatgga
25 agaggctacagaattagttcaagaatagagaataatggaccttcccaatgtttacatctgtgcccaggttgggtgaagacaagctga
aaattattatcctgaattactaaataatcttcatcagctaaatcacctcaacaaattttgtactgctagtaaaacttattatccttctatatctg
gtcttgacccaaagaatgtatttactgtaacagttatgccctgtacttcaaaaaatttgaagcagatagaccacaaatggaaaaagacg
gcctaagagatatagatgctgttataactactcgagaattagcaaaaatgattaaagatgctaaaataccatttgctaaactgaagatagc
gaagcagaccctgctatgggagaatacagcgggtgctgggtgccatattgggtgcaactggcggagttatggaagcagcttaagaagt
30 caaaagacttggctgaaaacgctgaactgaagatatagaatataagcaagttagaggattaaatggtataaaagaagctgaagtga
ataaaatacaacaaatataatgtagctgttataaatgggtgcttcaaatattttaagtttatgaaatctggtatgattaacgaaaaacaatatca
tttcatagaagtatggctgtcatggaggatgtgtaaatgtgggtggacagcccatgtaacccaaaagatttagaaaaagtagacat
aaaaaaagtaagagcttctgtattgtataatcaggatgaacatcttccaagagaaaatctcatgaaatactgcattagttaaaatgtatca
aaattatttggcaaacagggtgaaggctgtgccatgaaatattacactttaatatataaaaaataa

SEQ ID NO 175: *Desulfovibrio vulgaris* hydrogenase; Genbank accession number X02416

atgagccgtaccgtcatggagcgcatcgaatatgagatgcacactccggacccaaggccgatccggacaagctccacttcgtccag
atcgacgaggcaaaagtgcataggctgcgacacctgttcgagtactccccaccgcccgcattctcggcgaaatggggaaccgca
5 ctccattccccacatcgaggcgtgcatcaactgcggccagtgcctcacgcactgccccgagaacgccattctacgaggcacagtcgtg
gtgcctgaagtcgagaagaagctgaaggacggcaagggtgaatgcatcgccatgccccccccgcccgtgcgtatgactgggcg
acgccttcggcatgcccgtcggtccgtcaccaccggcaagatgctcgcggccctgcagaagctcggcttcgctcattgctgggaca
ccgagttcaccgctgacgtgacctctgggaagaggggtccgagttcgtggaacgcctcaccaagaagagcgacatgccgctgccg
cagttcacctcgtgctgccccggctggcagaagtatgccgagacctactacccgaactgctgccgcacttctccacgtgcaagtcgc
10 ccatcgccgatgaacggcgactggcgaagacctacggcgagagcggatgaagtacgacccaagcaggtctacaccgttccat
catgcccgtcatcgaaagaagtacgaagggttgcgtcccgaactgaagtcacgcgcatgcgcgacatcgacgccacgtgacca
cccgtagctggcctacatgatcaagaaggccggtatcgacttcgcaaacctcccacggcaagcgtgacagcctcatgggtgaat
ccaccggcggtgccaccatcttcggcgtcaccggcggtcatggaagcggcactccgcttcgctacgaagccgtcaccggcaa
gaagcccacagctgggacttcaaggcgtgcgcggtcttgatggcatcaaggaagccaccgtcaacgtcggcggtaccgacgtca
15 aggtcggcgtggtgcacggggccaagcgggtcaagcaggtctgcgacgatgtgaagcgggcaagtcgcctatcacttcacgaa
tcatgacctgccccggcggtcgtctgtggcgcggtcagcccgtcatgccccggtgctcgaagccatggaccgcaccaccac
ccgcctttacgcgggctgaagaagcgctcgcctatggcgagcgccaacaaggcatag

SEQ ID NO 176: *Entamoeba histolytica* hydrogenase; Genbank accession number

20 AF248542

atgccacctaaccatcacatacactcaccggacatgaccataaccatagatttcaatttgattggtctaaatgcatgggttggaatgt
gtgctactaaatgtacttttggggtgttagtaaaacaaccacaaaaattccaccatttgttcagcctaatagagaaaaactctcaagaa
aataccgacaagacaagagtacttattgatgagtcgaatgtactgggtgtggtcaatgttcttgggttgtaactttggttctattacccaa
tagaccatctgtgatacttttaaagctaaagaagctggaagaagctgttgctatgattgcacctcaactcgttaggtgtgtgctgagg
25 ctatgggaatgcctattggaagtacagctatggctcagttagttcattgtttaagacttattggattgattatgatttgatgttgctgga
gctgataagacaacaatggatgattatgccgaagttattgaaatgaaaaaaggaaaaggacctgctattacttcctgttgcctgctt
ggattgaacttgtgaaaaagaatatcctgacttaattccaacgtctctactgcccgttcaccaattggatgttagctgggtgtattaaaag
aggatgggcaaaggatgtagggaattgcagtagaagatcttactcgttggaataatgccttgattgctaaaaaacagagtcicaag
acaacaaattcatcaagactatgatgcttcatgtacttcaaatgaaattgctgcttatttcaaaaaacatctccacctgaagaatgtaaattt
30 acacaagaaagagaagaagcacttgtaaaactgaagatggtcaatgtgattaccatttagacgtatttctgggtgttctaataattttgga
aagactggaggagtttggaaactgtattgagagtaattgcacglaatgcaggagttgattggaacagttgactgttaacaaggaagaa
acttttaacatgctgcaagtggatcaacaatgacaaatcttctgttgatattggtggaactattatcacaggtgctgttgcacgtgtgtt
atgctattagacatgcttgaacttattgaaaaggaggttaaaagttgattgttgaaatgatggcatgtgttgagggttgccttgag
gagcaggtcaacaaaaattccaccagcaaaagaaacttgagatggataagagaagagtaattgtagatattttagatcaacaaactgat

attagagctgctaataaaaactgatgttcttgatggatgataaacatttgatcatcaaggcgacatcagcatcttcacacatattta
ctccagatatcaaaactaa

SEQ ID NO 177: *Scenedesmus obliquus* hydrogenase; Genbank accession number

5 AJ271546

atgcctgagtggaaccgggaggctggatgctgtttctgtccgcccgccagtgaacaggcgggctgtggggcagcagagcgag
gcgcctgtgtgctggggcagctggccaacagcagaatgtgattgccaccagctcccgcgccaaggccccgactggcagcag
acgctagatgagctagccaagcctaaggagcagcgcaaggatgatgcgccagatgcaccagcagtcgcgcgtggctattgcaga
gaccatgggactcaacctggggatgtgacagttggccagatggtagccggcctgcgcagctgggcttgattatgtttgacacgc
10 tgtttggctgctgacctcacatcatggaggaggcacagagctacggcacaggcttcaggaccacctggagcagcaccacaag
gaggagccgctgccatgttaccagctgctgccctggctgggtggcattggtggagaagtccaaccccgagctcatccctacctg
tcttcctgcaagtcgcccagatgatgtggcgagtcacaaactacttcgctgccgaggccggcgccaagcctgaggacatct
gcaacgtgagcgtgatgccctgcgtgcgaagcaggcgaggctgaccgcgagtggttaacaccacaggggctggcggcgcga
acgtggaccacgtcatgacaactgcagagctgggcaagatcttggtagcgcggaatcaagctgaacgacctgcaggagtcgcc
15 ttgacaaccccgctggcgaggcgaggcgggcgggcggtgctgttcggcaccactggaggcggtgatggaggcggcgctgcgcacc
gtgtacgaagtggcacacagaagcctttggaccgcatcgtcttggaggagtgccggcctggaggcgatcaaggagtccacgtg
cacctcacccaggccccaccagcccttaaggccttgaggcgagacggcaccggcatcacctcaacatcgcggtgccaa
cgccctcggaatgccaagaagctcatcaagcagctggctgcaggcgagagcaagtacgacttcagagtcagtcgctgccccg
gcggctgcatcgcgggcgggcgccagccgcgcagcggggacaagcagatcctgcagaagcgccaggcgggcattgacgacctg
20 gacgagcgcggggtatccggcgagccacgagaaccgctgattggcgcgctgtatgagaagttcctggggcgagcccaacggc
cacaaggcgacgagctgctgcacacgactacgtggccggcgggcggtgccgatgagaagtga

SEQ ID NO 178: *Chlorella fusca* hydrogenase; Genbank accession number AJ298227

atgtgttccccgtgggtgcaagtaggcacgcaggcggtgcaaggcatgttctgtccgtgcagcagggccaacatctgagtgatt
25 gtcctccaacacctcaggccaagctgcctcactggcagcaggctctggatgagctcgccaagccaaggagagcaggaggtgatg
atcgcgcaaatgccctccgctgttcgtgctgctattgctgagaccattggcttggcccaggagatgtaccattgggcagctcgtgact
gggctgcgtatgcttggcttgattatgtcttgacacctgtttggctgacctgaccattatggaggagggaacggagctgctgcatc
gcctgcaggacctctggagcagcaccacaaggaggagccactgccatgttcaccagttgctgccagggtgggttgccatg
gttgaaaagagcaatcctgagctcatccctacctgtcatcttgcaagtcgctcagatgatgcttggggccgtatcaagaactactatg
30 cacagcaggttgagtgagcccagtgacatctgcaacgtgtcagtcagtcacgcaagcaggagagggctgaccggga
gtgttcaacaccacaggtgcaggccttgccgctgattgatcatgtgtgactactgctgaggttgtaagatatcctggagcgtgg
catcaagctgaatgagctgccagagagcaacttgacaacccattggcgaggcgacaggtggctgctgctgtttggcaccactgga
gggtgcatggaggcagcacttcgcacagctatgaagtggtagccagaagcccatgggtcgtgttgactttgaggagggtgcgaggc
cttgaaggaaatcaaggaggcagagatcacactcaagccaggagacgacagccattcaaaccttcgagagctgatgggcagg

gcatcacgctcaagattgcagtagccaatgggcttgcaatgccaaagctcatcaagagcctgtcagagggaaggccaagtatg
atttcattgaggtcatggcatgccctgggtggctgcattggcggaggcggctagccccgcagtactgacaagcagatcctgcagaagc
gccagcaggctatgtacaacctggatgagcgcagtagcatccgccgcagccatgataaccattcatccaggcgctgtatgacaagtt
cctaggcgccaccaacagccacaaggcacatgatctgtgcacacacactatgtggcaggtggaattccagaggagaagtga

5

SEQ ID NO 179: Synthetic construct green fluorescent protein mRNA; Genbank accession
number AF188479

atggccaagggcgaggagctgttcaccgggtgtgtccccatcctgggtggagctggacggcgacgtgaacggccacaagttctccgt
ctccggcgagggtgaggggtgacgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgccctggc
10 ccacctggtcaccacctgacctacgggtgtgcagtgtcttctccgctaccccgaccacatgaagcagcagcacttctcaagtccgc
catgcccgagggtacgtgcaggagcgcacatcttctcaaggacgacggcaactacaagaccgcgccgagggtcaagttcgag
ggcgacacctggtgaaccgcatcgagctgaaggcgatcgacttaaggaggacggcaacatcctgggccacaagctggagtaca
actacaactcccacaacgtgtacatcatggccgacaagcagaagaacggcatcaaggtgaacttaagatccgccacaacatcgag
gacggctccgtgcagctggccgacctaccagcagaacacccccatcgcgatggccccgtgctgctgcccgacaaccactacct
15 gtccatccagtccgacctgtccaaggaccccaacgagaagcgcgaccacatggtctgtggtgagttcgtcaccgctgccggcatcac
ccacggcatggacgagctgtacaagtaa

SEQ ID NO 180: Nonshuffled segment IX:

atccgtagttatccttatggccatcttagcgcagttgggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggcccaacaag
20 tcgctcgcggtcgctgtcggcgccaaacccgcagctgcatccaccagattcacttgttagatcgacctaggttgcgggaccggagg
cggctcgctgtgcaagcgcggtgacctgtacggcgcatggatgccatctcgattcgcgcggcagaatcgggccccgcgcacat
ttaagccgcgggcgagactcatttcgttacgtgcatcgattaacagcttctggacctga

SEQ ID NO 181: Nonshuffled segment X:

ttaaagtcgtacgtccaagtataactaagccgacgtcgaccactctagaggatcgatccccgctccgtgtaaattggaggcgctcgtt
25 gatctgagccttggccccctgacgaacggcggtggatggaagatactgctctcaagtgtgaagcggtagcttagctccccgttctgctg
tgatcagctcttttcaacacgtaaaaagcggaggagttttgcaattttgttggtgtaacgatcctccgttgattttggcctcttctccatggg
cgggctgggctatttggcagttgggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggcccaacaagtcgcctcgcg
gtcgtgtcggcgccaaacccgcagctgcatccaccagattcacttgttagatcgacctaggttgcgggaccggaggcggctcgctgt
30 gcaagcgcgggtgacctgtacggcgcatggatgccatctcgattcgcgcggcagaatcgggccccgcgcacatttaagccgcg
ggcgagactcatttcgttaaattctgatacatgctattcagatcttaca

SEQ ID NO 182: Nonshuffled segment XI:

tcttccatcgtaaaictagcatcgattagcccgacgtcgaccactctagaggatcgatccccgtccgtgtaaatggaggcgctcggtg
atctgagccttgccccctgacgaacggcggtggatggaagatactgctctcaagtgtgaagcggtagcttagctccccgttctgct
gatcagtccttttcaacacgtaaaaagcggaggagttttgcaattttgttggtgtaacgatcctccgttgattttggcctcttctccatgggc
gggctgggcgtatttgagtggtcaggggctggcgacgcgctgctgacgcgcaagtgaatggcccaacaagtcgcctcgcggt
5 cgctgtcggcgccaaaccgcagctgcatccaccagattcacttgtagatcgacctagggtgcgggaccggaggcggtcgctgtg
caagcgcggtgacctcgtacggcgcatggatcgccatctcgattcgcgggcagaatcgggccccgcgcacatttaagccgcgg
gcgagactcattcgtaatctgtaataatctagtcgaggcattcaag

SEQ ID NO 183: Nonshuffled segment XII:

10 atctgtaataatctagtcgaggcattcaagatggccaaggcgaggagctgtcaccgggtgtggtcccatcctggtggagctggacg
gcgacgtgaacggccacaagttctccgtctcggcgagggtgagggtagcggccacctacggcaagctgacctgaagttcatctgc
accaccggcaagctgcccgtgccctggcccaccctgggtcaccacctgacctacgggtgtgcagtgtctctccgctacccccgaccac
atgaagcagcacgacttctcaagtccgccatgcccaggggctacgtgcaggagcgcaccatcttctcaaggacgacggcaactac
aagacccgcgcccagggtcaagttcagggcgacaccctggtgaaccgcatcgagctgaagggcacgacticaaggaggacggc
15 aacatcctgggccacaagctggagtacaactacaactcccacaacgtgtacatcatggccgacaagcagaagaacggcatcaaggt
gaacttaagatccgccacaacatcgaggacggctccgtgcagctggccgaccactaccagcagaacacccccatcgcgcatggc
cccgctgtctgcccgaaccactacctgtccatccagtcgcctgtccaaggaccccaacgagaagcgcgaccacatggctctg
ctggagttcgtcaccgtgccggcatcaccacggcatggacgagctgtacaagtaaaactggcttaaatcgtaaacatcgtgtga

20 SEQ ID NO 184: Nonshuffled segment XIII:

aactggcttaaatcgtaaacatcgtgtgaccgacgtcgaccactctagaggatcgatccccgtccgtgtaaatggaggcgctcggt
gatctgagccttgccccctgacgaacggcggtgatggaagatactgctctcaagtgtgaagcggtagcttagctccccgttctgctg
tgatcagtccttttcaacacgtaaaaagcggaggagttttgcaattttgttggtgtaacgatcctccgttgattttggcctcttctccatggg
cgggctgggcgtatttggaattaacataactgtcgattaccgtgcga